REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
3 Jun 96

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
Bounding The Edge Cover Time Of Random Walks On Graphs

5. FUNDING NUMBERS

6. AUTHOR(S)

Eric Richard Bussian

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

AFIT Student Attending:

Georgia Institute of Tech

8. PERFORMING ORGANIZATION REPORT NUMBER

96-021D

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DEPARTMENT OF THE AIR FORCE
AFIT/CI
2950 P STEET, BLDG 125
WRIGHT-PATTERSON AFB OH 45433-7765

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release IAW 190-1
Distribution Unlimited
BRIAN D. GAUTHIER, MSgt, USAF
Chief Administration

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

19960809 063

14. SUBJECT TERMS

15. NUMBER OF PAGES
102

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT

18. SECURITY CLASSIFICATION OF THIS PAGE

19. SECURITY CLASSIFICATION OF ABSTRACT

20. LIMITATION OF ABSTRACT

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	**PR**	- Project
G	- Grant	**TA**	- Task
PE	- Program Element	**WU**	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE - See authorities.
NASA - See Handbook NHB 2200.2.
NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - Leave blank.
DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA - Leave blank.
NTIS - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

★U.S.GPO:1990-0-273-271

Bounding The Edge Cover Time Of Random Walks On Graphs

Eric Richard Bussian
Major, USAF, 1996
Doctor of Philosophy
Georgia Institute of Technology

102 Pages

Directed by Dr. Richard A. Duke

In recent years a great deal of attention has been focused on answering questions regarding the cover times of random walks on simple graphs. In this dissertation we answer questions about the edge cover time of such walks.

We begin by reviewing many of the definitions and established results for vertex cover time. We present the little that has been published regarding bounds on the edge cover time. Having completed this review we establish a new, and sometimes more useful, global upper bound for edge cover time.

We then narrow our focus and consider the edge cover time of the path. We establish an exact description of the edge cover time for a random walk on the path started at an endpoint in terms of coefficients related to the Bernoulli Numbers of the Second Kind. Studying these coefficients carefully allows us to develop a tight bound on this cover time of $(n-1)^2+\Theta(n^2/\log n)$. Using these results, and generalizing, provides a description of the edge cover

time for walks on the path started from an arbitrary vertex. This generalization gives us a bound of $(5/4)(n-1)^2 + O(n^2/\log n)$ for the edge cover time for the path.

Having established a tight bound for walks on paths we then focus on other trees. We prove that the edge cover time for a random walk started from the center of a star graph minimizes edge cover time for walks on all trees on n vertices. We also establish the fact that in all graphs the edge cover time for a walk started from a leaf is always greater than for a walk started from its point of attachment. We continue our study of trees by establishing a global upper bound on the edge cover time for all trees and use it to study balanced k-ary trees.

Finally we show the connection between our previous developments for the edge cover time for paths and that of the edge cover time on the cycle. This leads to a tight bound for the undirected and directed edge cover times of this graph.

Bibliography

[1] D.J. Aldous. "On the Time Taken by Random Walks on Finite Groups to Visit Every State". *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, pp. 361-374, **62** (1983).

[2] D.J. Aldous. "Reversible Markov Chains and Random Walks on Graphs". Draft of first six chapters of book, January 26, 1993.

[3] R. Aleliunas, R. M. Karp, R. J. Lipton, L. Lovász, and C. Rackoff. "Random Walks, universal traversal sequences, and the complexity of maze problems". *20th Annual Symposium on Foundations of Computer Science*, pp. 218-223, San Juan, Puerto Rico, October 1979.

[4] G. Barnes and U. Fiege. "Short Random Walks on Graphs". *Proceedings of the 25th Annual ACM Symposium on Theory of Computing*, pp. 728-737, May 1993.

[5] J.A. Bondy and U.S.R. Murty. *Graph Theory with Applications.* New York: Elsevier North Holland, Inc., 1976.

[6] G. Brightwell and P. Winkler. "Maximum hitting times for random walks on graphs". *Random Structures and Algorithms*, 1:263-276, 1990.

[7] G. Brightwell and P. Winkler. "Extremal Cover Times for Random Walks on Trees". *Journal of Graph Theory*, 14:5, pp. 547-554(1990).

[8] Paul S. Bruckman. private communication.

[9] A. Chandra, P. Raghavan, W. Ruzzo, R. Smolensky and P. Tiwari. "The Electrical Resistance of a Graph Captures its Commute and Cover Times". *Proceedings of the 21st Annual ACM Symposium on Theory of Computing*, pp. 574-586, May 1989.

[10] D. Coppersmith, U. Feige and J. Shearer. "Random Walks on Regular and Irregular Graphs". *Preprint*, September 22, 1993.

[11] P.G. Doyle and J.L. Snell. *Random Walks and Electrical Networks*, The Mathematical Association of America, 1984.

[12] William Feller. *An Introduction to Probability Theory and Its Applications*, Vol 1, John Wiley and Sons, Inc. 1957.

[13] Uriel Fiege. "A Tight Upper Bound on the Cover Time for Random Walks on Graphs". *Random Structures and Algorithms*, Vol 6, No. 1, 1995.

[14] Uriel Fiege. "A Tight Lower Bound on the Cover Time for Random Walks on Graphs". *Preprint*, October 21, 1993.

[15] Daniel C. Fielder. private communication.

[16] F. T. Howard. "Congruences and Recurrences for Bernoulli Numbers of Higher Order". *The Fibonacci Quarterly*, Vol 32, No. 4, pp. 316-328, 1994.

[17] Charles Jordan. *Calculus of Finite Differences.* New York: Chelsea, 1965.

[18] J.G. Kemeny and J.L. Snell. *Finite Markov Chains*, Springer-Verlag, 1983.

[19] László Lovász and Peter Winkler. "Note: On The Last New Vertex Visited by a Random Walk". *DIMACS Technical Report*, 91-38, May 1991.

[20] Peter Matthews. "Covering Problems for Brownian Motion of Spheres". *The Annals of Probability*, Vol 16, No. 1, pp. 189-199, 1988.

[21] Niels Erik Nörlund. *Vorlesungen Über Differenzenrechnung.* New York, Chelsea, 1954.

[22] Prasad Tetali. "Random Walks and the Effective Resistance of Networks". *Journal of Theoretical Probability*, 4:101-109, 1991.

[23] Herbert S. Wilf. "The Editor's Corner: The White Screen Problem". *MAA Monthly*, Vol 96, No 8, pp. 704-707, Oct '89.

[24] P. Winkler. "Extremal Problems for Random Walks on Graphs". *Preprint.*

[25] P. Winkler and D. Zuckerman. "Multiple Cover Time". Unpublished Extended Abstract.

[26] David Zuckerman. "Covering Times of Random Walks on Bounded Degree Trees and Other Graphs". *The Journal of Theoretical Probability,* Vol 2, No 1, pp. 147-157, Jan '89.

[27] David Zuckerman. "On the time to traverse all edges of a graph". *Information Processing Letters*, 38:335-337, 1991.

[28] David Zuckerman. "A Technique for Lower Bounding the Cover Time". *SIAM Journal of Discrete Mathematics*, Vol. 5, No. 1, pp. 81-87, Feb '92.

ABSTRACT ON
OTHER SIDE

A DISSERTATION
Presented to
The Faculty of the Division of Graduate Studies

by

Eric R. Bussian

In Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy in Mathematics

Georgia Institute of Technology
June 1996

BOUNDING THE EDGE COVER TIME OF RANDOM WALKS ON GRAPHS

Approved:

R. A. Duke

Richard A. Duke

Neil J Calkin

Neil J. Calkin

Prasad Tetali

Ronald W. Shonkwiler

R. Gary Parker

Date Approved by Chairman JUNE 3, 1996

Dedication

I dedicate this, my first published work, to my parents, Richard and Margaret Bussian. Without their love, support and encouragement I would not have accomplished this feat. The seeds of faith in Christ that they planted in me have provided the strength I have needed. Thank you Mom and Dad.

Acknowledgements

As I reflect back on the years of effort expended to accomplish this task I see the supporting hands and spirits of many people in my life. I wish to especially thank the following.

First, thanks to my primary advisor Richard Duke. His patience and careful monitoring of my progress helped ensure I reached this goal on schedule. Thanks also to my other defense committee members: to R. Gary Parker, ISYE, who served as an alternate advisor throughout the last three years, to Neil Calkin and Prasad Tetali who shared their insight and intuition, and to Ronald Shonkwiler who juggled his already overloaded schedule to provide his expertise.

I owe a great debt to James Herod who has been an advisor, confidant and friend since I first arrived at Tech to earn my master's degree in 1988. His advise and encouragement helped me through the toughest days. My thanks also to Jack Howley whose support and counsel have been instrumental in bringing me to this point in my life.

I would not have had the opportunity to pursue this degree if it were not for the support of Colonel Daniel Litwhiler and Colonel Steven Gordon. Col. Litwhiler's faith in me as I struggled through this process was a source of strength and inspiration.

Tim Howard was there every step of the way supporting and encouraging me even as he went through this same process. It was he who was able to most clearly empathize with my dilemmas and frustrations. His patience and counsel were instrumental in the success of this work. Thanks Tim.

Thanks to Greg Watson, Kate Klee and Tom Fowler, all of whom have helped me in my efforts while struggling with their own work. And to Eric and Carol Johnson who have been there for me in so many ways in the last three years.

Finally my thanks to God and the many friends and supporters he provided through church and volunteer activities. To Ginger and Stewart Cashin, Michele Greenwell, Gwen Hughes, Danny Ingram, David Yackley and Father Alan Dillman, thanks. All of these individuals have enriched my life and provided the strength to make this task a success.

Contents

List of Tables

List of Figures

Table of Symbols

B_ϵ	Blanket time for a graph
C_n	The cycle on n vertices
C^{ue}	Undirected edge cover time... $\max_x E_x C^{ue}$
$C_{x,y}$	Commute time between vertices x and y ... $E_x H_y + E_y H_x$
$d(x)$	Degree of the vertex x
$d_-(x)$	In a graph with directed edges, the indegree of x
$d_+(x)$	In a graph with directed edges, the outdegree of x
C^e	Directed edge cover time... $\max_x E_x C^e$
C^v	Vertex cover time... $\max_x E_x C^v$
$E_s \bullet$	Expected value of $\bullet$ starting at s
$E_x C^e$	Expected number of steps to traverse all directed edges at least once starting from vertex x
$E_x C^{ue}$	Expected number of steps to traverse all undirected edges at least once starting from vertex x
$E_x C^v$	Expected number of steps to visit all vertices at least once starting from vertex x
$E_x H_y$	Expected number of steps to walk from x to y
γ	Euler's Constant, approximately 0.577216
H	Maximum hitting time... $\max_{x,y} E_x H_y$
h_n	n^{th} Harmonic Number ... $\sum_{k=1}^{n} \frac{1}{k}$
$N_x(t)$	Number of times a random walk has visited x by time t
π	Stationary distribution of a random walk
P_n	The path on n vertices
S_n	The star on n vertices

Table 0.1: Symbols

Summary

In recent years a great deal of attention has been focused on answering questions regarding the cover times of random walks on simple graphs. In this dissertation we answer questions about the edge cover time of such walks.

We begin by reviewing many of the definitions and established results for vertex cover time. We present the little that has been published regarding bounds on the edge cover time. Having completed this review we establish a new, and sometimes more useful, global upper bound for edge cover time.

We then narrow our focus and consider the edge cover time of the path. We establish an exact description of the edge cover time for a random walk on the path started at an endpoint in terms of coefficients related to the Bernoulli Numbers of the Second Kind. Studying these coefficients carefully allows us to develop a tight bound on this cover time of $(n-1)^2+\Theta(n^2/\log n)$. Using these results, and generalizing, provides a description of the edge cover time for walks on the path started from an arbitrary vertex. This generalization gives us a bound of $(5/4)(n-1)^2 + O(n^2/\log n)$ for the edge cover time for the path.

Having established a tight bound for walks on paths we then focus on other trees. We prove that the edge cover time for a random walk started from the center of a star graph minimizes edge cover time for walks on all trees on n vertices, paralleling a result for vertex cover time on trees. In the process of proving this we also establish the fact that in all graphs the edge cover time for a walk started from a leaf is always greater than for a walk started from its point of attachment. We continue our study of trees by establishing a global upper bound on the edge cover time for all trees and use it to study balanced k-ary trees. Our study of the star also leads to alternate proofs for two identities involving the Stirling Numbers of the Second Kind.

Finally we show the connection between our previous developments for the edge cover time for paths and that of the edge cover time on the cycle. This leads to a tight bound for the undirected and directed edge cover times of this graph.

CHAPTER I

DEFINITIONS AND KNOWN RESULTS FOR VERTEX COVER TIME

1.1 Definitions

A *graph*, G, consists of a pair of sets $\{V, E\}$. The elements of the set V are known as vertices. The elements of E are known as edges and consist of unordered pairs of the elements of V. In this thesis we will consider only finite simple graphs, so V will be a finite set, and E will contain only distinct pairs from V in which each element of the pair is also distinct. We will follow common convention and normally indicate the size of the vertex set by n and the size of the edge set by m. The *neighbors* of a vertex x are the elements of the set $\{y : \exists\{x, y\} \in E\}$. The *degree* of a vertex $x, d(x)$, is the cardinality of the neighbor set of x. We will also discuss the *directed* version of G in which we replace each element, $\{x, y\}$, of the set E by two ordered pairs,

(x, y) and (y, x). In this case we may make use of the *indegree*, $d_-(x)$ or *outdegree*, $d_+(x)$, of a vertex x. However, due to the method of construction of the graph it is clear that $d_+(x) = d_-(x) = d(x)$. We will rely on the usual definitions of other graph concepts such as whether a graph is connected, bipartite or acyclic, etc. A standard reference for these concepts is Bondy and Murty[5].

A *discrete Markov chain* is a stochastic process in which

$$P\{X_n | X_{n-1}, X_{n-2}, \dots, X_0\} = P\{X_n | X_{n-1}\},$$

and this probability is independent of n. The set of possible outcomes for X_n is known as the *state space* of the chain. We will denote by $p_{i,j}$ the probability $P\{X_n = j | X_{n-1} = i\}$. If $P = [p_{i,j}]$ is the state transition probability matrix for an n state discrete Markov chain, the chain is *irreducible* if for each pair, $\{i, j\}$, of elements of the state space there exists some integer, $k_{\{i,j\}}$, such that the i, j entry of $P^{k_{\{i,j\}}}$ is nonzero. The probability vector, π, is the *stationary distribution* for an irreducible Markov chain if $\forall j, \pi_j \geq 0$, $\sum_j \pi_j = 1$, and $\pi P = \pi$. A Markov chain is reversible if for all i and j in the state space $\pi_i p_{i,j} = \pi_j p_{j,i}$.

A *random walk* on a connected graph $G = \{V, E\}$ is a reversible, irreducible Markov chain with state space V and for which $p_{x,y} = 1/d(x)$ when there is an edge $\{x, y\}$ in E and $p_{x,y} = 0$ otherwise. One may think of this as placing a token on a vertex of the graph and then observing it as at each time step it moves uniformly at random to some neighbor of its present position. This definition may be generalized to cases in which the transition probabilities at each vertex are not uniform, but we will not address this case. The classical example of a random walk on a graph is the "gambler's ruin problem." If one considers a fair game, i.e. one in which the probabilities of success and failure are the same, then gambler's ruin can be viewed as placing a token at some nonnegative point on the natural number line and asking how long before the token will reach either 0 or some predetermined upper limit when each play of the game results in a gain or loss of 1 unit.

Throughout this dissertation we will discuss random variables whose value may depend on the starting point of the walk. For a random variable X, we will denote the expected value of X for a walk started at vertex z by E_zX. The *hitting time from x to y*, E_xH_y, is the expected number of steps it will take for the token to move from vertex x to vertex y. Note that it is not generally the case that $E_xH_y = E_yH_x$. We will denote by H the maximum hitting time for all pairs of vertices, i.e. $H = max_{x,y \in V} E_xH_y$. The commute time between vertices x and y is $C_{x,y} = E_xH_y + E_yH_x$

Given a starting vertex x, and counting this placement as a visit to x, one may ask: What is the expected number of steps needed to ensure the random walk has visited all the vertices of the graph at least once? The *vertex cover time*, C^v, is the maximum of this value over all starting points x. One may modify the question slightly and ask: What is the expected number of steps to visit all the vertices and then return to the starting point? This time is called the *vertex cover and return time.*

We will also use the following notation to describe the growth of a function. When f and g are nonnegative valued functions we define $f(n) = O(g(n))$ if $\exists$ N and c, $\ni \forall n \geq N, f(n) \leq cg(n)$. Likewise $f(n) = \Omega(g(n))$ if $\exists$ N and c, $\ni \forall n \geq N, f(n) \geq cg(n)$. Define $f(n) = \Theta(g(n))$ if $f(n) = \Omega(g(n))$ and $f(n) = O(g(n))$. Finally $f(n) = o(g(n))$ if $\lim_{n\to\infty} f(n)/g(n) = 0$.

1.2 Results for Vertex Cover Time

In recent years a good deal of research has been devoted to questions regarding the hitting and vertex cover times of random walks on graphs. Unfortunately, there are only a few cases where the quantities are easily computed exactly. The following well known recursive argument provides the hitting time from one endpoint to the other of a path with n vertices.

Assume that the vertices of the path are labeled in the natural way from 1 to n. Define a_j to be the expected number of steps needed, starting at vertex j, to first reach vertex n. Clearly $a_n = 0$. Then since a walk started at vertex 1 steps to vertex 2 with probability 1 on its first step we must have

$$a_1 = 1 + a_2.$$

Finally for $1 < j < n$ we have

$$a_j = \frac{1}{2}(1 + a_{j-1}) + \frac{1}{2}(1 + a_{j+1}).$$

Now we have a second order recursion with two boundary conditions whose solution provides the closed form description

$$a_j = (n-1)^2 - (j-1)^2.$$

Thus $E_1 H_n = (n-1)^2$.

Using a similar recursive argument one may establish that the vertex cover time for a walk started at vertex j is $(5/4)(n-1)^2 - (1/4)(n-2j+1)^2$. This quantity is maximized by a walk started at the center/centers of the path, and has value $(5/4)(n-1)^2$ there when n is odd. Another graph for which a recursive argument provides the vertex cover time is the cycle, resulting in $\binom{n}{2}$ for the n vertex cycle.

A further simple graph one may consider is the *Star*, S_n, consisting of $n-1$ vertices of degree 1 and a single vertex of degree $n-1$. In this case a "coupon collector" type argument provides the exact vertex cover time.

Consider a walk on S_n started at a leaf. Then one may consider the leaves of the star to be coupons. On every second step a coupon will be chosen. Upon first acquiring the j^{th} new coupon consider how many steps, on average, one must take to acquire the $(j+1)^{st}$. Having acquired j distinct coupons, with probability $(j-1)/(n-1)$ one takes two steps and returns to a leaf already visited, otherwise, with probability $(n-j-1)/(n-1)$ one takes two steps and acquires a new coupon. Thus for $j \geq 1$, if c_j represents the number of steps to acquire the $(j+1)^{st}$ coupon, we have

$$c_j = \frac{j}{n-1}(2+c_j) + \frac{n-j-1}{n-1}(2),$$

which has solution

$$c_j = 2\left(\frac{n-1}{n-j-1}\right).$$

Then

$$E_{leaf}C^v = 2(n-1)\sum_{j=1}^{n-2}\left(\frac{1}{n-j-1}\right) = 2(n-1)h_{n-2},$$

where $h_k = \sum\limits_{j=1}^{k}(1/j)$, the k^{th} harmonic number. It is easy to see that if a walk starts from the center of S_n it would take one step to reach a leaf and then accomplish a vertex cover as if it had started at that leaf. Thus for S_n,

$$C^v = 2(n-1)h_{n-2} + 1.$$

Brightwell and Winkler[7] established that the vertex cover time for all trees on n vertices is minimized by the star. They conjectured that a random walk on the path when started from the center/centers maximizes vertex cover time for trees, but this has yet to be proved.

Modifying the "coupon collector" argument above only slightly results in an easy proof that the vertex cover time for the clique on n vertices is $(n-1)h_{n-1}$.

Brightwell and Winkler[6] determined the graph which maximizes hitting time among all graphs. This graph is known as the "lollipop" and consists of a clique on approximately $2n/3$ vertices attached to a path on the remaining vertices. They proved that the hitting time in this graph is maximized by starting at a vertex in the clique and proceeding to the tail of the path. The hitting time for such a walk is $(4/27 + o(1))n^3$. They conjectured that this graph also maximized vertex cover time for all graphs. Fiege[13] has proved that the maximum vertex cover time for all graphs is bounded above by $(4/27)n^3 + O(n^{2.5})$ lending credence to that conjecture.

Since there are few graphs for which it has been possible to compute hitting or vertex cover times exactly, a good deal of work has been done to establish bounds on these quantities. One of the earliest bounds for vertex cover time is a consequence of a more general theorem due to Matthews[20]. This result provides an upper bound on the vertex cover time based on the maximum hitting time of the graph.

Theorem 1.1 *If G has n vertices then*

$$C^v \leq H h_n \sim H \log n \tag{1.1}$$

Matthews developed this result using Markov chain theory while studying Brownian motion on the sphere. While it proves to be too generous in some cases there are a number of "natural" cases in which it is tight. One such example is the hypercube for which Aldous[1] proved a cover time bound of $\Theta(n \log n)$. Another is the balanced k-ary tree for which Zuckerman proved $C^v = \Theta(n \log^2 n)$.

Doyle and Snell[11] make an elegant connection between the theory of resistor/capacitor networks and random walks on graphs. Building on this foundation Chandra[9] et al. prove that the cover time for d-regular expander graphs is $\Theta(n \log n)$. They also use this theory to establish bounds of $O(n \log^2 n)$ for 2 dimensional toruses and $O(n \log n)$ for d dimensional toruses where $d > 2$ and $n = k^d$ for some integer k. Matthews bound is tight in all these cases as well.

While attempting to prove the existence of universal traversal sequences Aleliunas[3] et al. established perhaps the first general upper bound on the vertex cover time.

Theorem 1.2

$$C^v \leq 2m(n-1). \tag{1.2}$$

Here, as usual, n denotes the number of vertices in the graph, and m the number of edges. Their technique made use of the fact that the commute time between any pair of adjacent vertices in a graph is bounded by $2m$. They observed that if one takes any spanning tree of a graph and any specific vertex x, there is a covering walk which begins and ends at x and traverses each edge of the spanning tree exactly once in each direction. Then the vertex cover time is bounded above by the average time for a walk to cover all the vertices of the graph in the order imposed by this walk. The amount of time necessary for such a walk is just the sum of the commute times for each pair of adjacent vertices in the spanning tree which is bounded by $2m(n-1)$.

A trivial lower bound for the vertex cover time of all graphs is n. Fiege[14] has proven the following lower bound which is tight for several classes of graphs.

$$C^v \geq n \log n. \tag{1.3}$$

Other authors have proved bounds on the vertex cover time for specific classes of graphs using the properties of regularity, the average or maximum degree, and the diameter of the graph. Aldous has collected these, and much more about reversible Markov chains, in several chapters of a book[2] yet to be published.

CHAPTER II

A NEW BOUND ON THE EDGE COVER TIME

2.1 Definitions and Known Results

Given the interest in vertex cover time it is only natural to ask the question in terms of edges as well. That is, given a starting point x, what is the expected number of steps necessary to ensure that the token has traversed every edge of G at least once. The *undirected edge cover time*, C^{ue}, is the maximum over all starting points of this value. The *directed edge cover time*, C^{e}, is the maximum over all starting points for this value when one asks the question in terms of traversing each edge of G in both directions. In this work we will concentrate on answering some of the questions related to these quantities.

There appears to be relatively little in the literature about the edge cover time. One of the few results is a global bound due to Zuckerman[27].

Theorem 2.1 *The directed edge cover time for a graph G is $O(nm)$.*

In his paper Zuckerman actually establishes two upper bounds. The first is the "natural" bound derived from Matthews[20] results. It is easy to argue that the maximum hitting time from one directed edge to another in a graph is bounded above by $(2m + H)$. Thus, using Matthews' result, and the fact that the sequence of directed edges traversed during a random walk also forms a Markov chain, one obtains $C^e = O((m + H) \log m)$. Zuckerman points out that this bound is tight for those graphs having $C^v = \Theta(H \log n)$, such as cliques, hypercubes, and k-dimensional toruses. However, in the general case this natural bound yields $O(nm \log n)$. The path is an example where this fails to be tight, for it is clear that the edge cover time for a path is bounded by $2n^2$, while this bound provides only $O(n^2 \log n)$.

In order to remove the extra $\log n$ factor in the "natural" bound Zuckerman makes use of a renewal type argument to show that $C^e = O(k(C^v + m^{1+1/k}))$, for any integer k. Letting $k = 2$ here provides the $O(mn)$ bound. This yields bounds of $O(n^2)$ for the path and $O(n^3)$ for the "lollipop," which are tight. However, this also yields $O(n^\alpha)$, with $\alpha > 2$, for the clique. This exceeds the "natural" bound of $O(n^2 \log n)$, which is also tight. The fact that neither of these bounds is tight in all known cases led us to search for another means to bound the edge cover time.

2.2 A New Global Upper Bound

Intuitively, if one focuses on the edges emanating from a specified vertex x, then the amount of time necessary to traverse all the edges out of x may be found by a "coupon collector" argument. It is well known that the expected amount of time to collect all of the coupons in a set of size n is approximately $n \log n$. Thus it seems plausible that if we visit vertex x a constant multiple of $d(x) \log d(x)$ times, then with high probability we should have traversed all the edges out of x. Now if we consider a regular graph of degree d with n vertices and $m = dn$ directed edges, and *if* we were near stationarity, then, since the stationary distribution has $\pi_j = d/m, \forall j$, the number of steps, k, needed would be given by

$$k\left(\frac{d}{m}\right) \approx d \log d \Rightarrow k \approx m \log d.$$

Since $d < n$ this would indicate that $k \approx dn \log n$ should certainly suffice in such cases. The question becomes: Is there a way to quantify, in the general case, how many times one has visited each vertex on a "long" walk through the graph?

We know that eventually the frequency with which each vertex is visited approaches the value determined by the stationary distribution. Winkler and Zuckerman[25] have defined the *blanket time* of a graph, B_ϵ, to be the expected time for a random walk to have hit every vertex of G within a multiplicative constant factor, $\epsilon < 1$, of the number of times predicted by the stationary distribution. Specifically, letting $N_x(t)$ be the number of times vertex x is visited by time t, they define $\mathbf{B}_\epsilon = \min\{t : (\forall x) N_x(t) \geq (1-\epsilon)\pi_x t\}$, and set $B_\epsilon = \max_{x \in V} E_x \mathbf{B}_\epsilon$. Clearly $B_\epsilon = \Omega(C^v)$. They conjecture that $B_\epsilon = O(C^v)$. They prove their conjecture in the cases of the path, the cycle, and those graphs for which $C^v = \theta(H \log n)$.

Using the concept of Blanket Time and letting $\bar{d} = \left(\sum_{x \in V} d(x)\right)/n$, the average degree of a vertex in the graph, we now claim:

Theorem 2.2

$$C^e = O(\bar{d} B_\epsilon). \tag{2.1}$$

Proof

Since B_ϵ is an expected value, we know from Markov's Inequality that $\forall c > 0$,

$$P\{\exists x \in V \ni N_x(c\bar{d}B_\epsilon) < (1-\epsilon)\,\pi_x c\bar{d}B_\epsilon\} \leq \frac{1}{c\bar{d}}. \tag{2.2}$$

Thus with probability $1 - 1/c\bar{d}$ we have $\forall x \in V, N_x(c\bar{d}B_\epsilon) \geq (1-\epsilon)\,\pi_x c\bar{d}B_\epsilon$.

On each visit to a vertex x the probability that a specified directed edge, say (x, y), is not traversed on the next step of the walk is

$$\Pr[\text{edge } (x,y) \text{ is not traversed}] = \left(1 - \frac{1}{d(x)}\right) < e^{\frac{-1}{d(x)}}.$$

Therefore

$$\Pr[\text{edge } (x,y) \text{ is not traversed after } q \text{ visits to } x] < e^{\frac{-q}{d(x)}}.$$

So in those cases where, after taking $q = c\bar{d}B_\epsilon$ steps(c to be chosen later), we have "blanketed" we have

$$\Pr[\text{edge } (x,y) \text{ is not traversed after } c\bar{d}B_\epsilon \text{ steps}] < e^{\frac{-(1-\epsilon)\pi_x c\bar{d}B_\epsilon}{d(x)}}.$$

Using the fact that $\pi_x = d(x)/m$ we have

$$\Pr[\text{edge } (x,y) \text{ is not traversed after } c\bar{d}B_\epsilon \text{ steps}] < e^{\frac{-(1-\epsilon)c\bar{d}B_\epsilon}{m}}.$$

Note that this value is independent of the particular edge we are considering so

$$\Pr[\exists \text{ an edge } (x,y) \text{ which is not traversed after } c\bar{d}B_\epsilon \text{ steps}] < me^{\frac{-(1-\epsilon)c\bar{d}B_\epsilon}{m}}.$$

Using the fact that $m = \bar{d}n$ we have

$$\Pr[\exists \text{ an edge } (x,y) \text{ which is not traversed after } c\bar{d}B_\epsilon \text{ steps}] < \bar{d}ne^{\frac{-(1-\epsilon)cB_\epsilon}{n}}.$$

Combining this with $B_\epsilon \geq C^v$ and inequality(1.3) yields

$$\Pr[\exists \text{ an edge } (x,y) \text{ which is not traversed after } c\bar{d}B_\epsilon \text{ steps}] < \bar{d}n^{1-(1-\epsilon)c}.$$

Then since $\bar{d} < n$, and if we restrict $\epsilon \leq (1/2)$, we observe that choosing $c \geq 6$ provides an upper bound of $1/n$ on this probability. If we restrict ourselves to considering graphs on $n \geq 12$ vertices, and combine this probability with that in equation(2.2), we may now claim that

$$\Pr[\text{we have not traversed all edges after } 12\bar{d}B_\epsilon \text{ steps}] < \frac{1}{6}.$$

Using the independence properties of random walks we may then say

$$\Pr[\text{Number of steps to complete an edge cover} > j12\bar{d}B_\epsilon \text{ steps}] < \left(\frac{1}{6}\right)^j.$$

This fact provides the result in our theorem. ∎

If Winkler and Zuckerman's conjecture that $B_\epsilon = O(C^v)$ is correct, then we have the new general bound for the edge cover time, $O(\bar{d}C^v)$. Unfortunately this bound proves to overestimate the order of the edge cover time on the lollipop. Winkler and Zuckerman have proved however, that the "blanket" time is $O(C^v)$ in all the "natural" cases such as cliques, hypercubes and expanders, as well as for the path and cycle. Thus we may use $O(\bar{d}C^v)$ in those cases, which turns out to be an improvement over using either of Zuckerman's bounds in isolation.

CHAPTER III

THE DIRECTED EDGE COVER TIME OF A WALK ON THE PATH

3.1 Starting from an Endpoint

It was mentioned in Chapter 1 that one of the classical examples of a random walk is the "Gambler's Ruin Problem," a probabilistic game played on the path. We also mentioned that the path is conjectured to be the graph which maximizes vertex cover time for walks on trees. The path also plays a central role in questions concerning directed edge cover times. In this chapter we determine the directed edge cover time for a walk begun at an endpoint of a path. We also obtain a bound on the directed edge cover time for a walk begun at any point of the path and show that, as for the vertex cover time, this quantity is maximized for a walk begun at the center of the path, when it is long enough. (Somewhat surprisingly, however, the directed edge cover time is not minimized by starting at an endpoint.)

We will use P_n to denote the path on the integers $[1 \ldots n]$. That is the tree on the vertex set $[1 \ldots n]$ in which for $1 < j < n$ there are edges $\{j-1, j\}$ and $\{j, j+1\}$. We will first investigate the edge cover time for a random walk on P_n which starts from an endpoint. We will consider the endpoint at which the walk starts to be labeled vertex 1 and the other endpoint to be vertex n.

It is obvious that the undirected edge cover time in this case, and for all trees, is the same as the vertex cover time. Here we wish to compute the expected directed edge cover time for this walk. After a little thought it becomes apparent that prior to completing a directed edge cover a walk will first complete a vertex cover of the path, and that the vertex cover portion of the walk will end upon its first visit to vertex n. From that point the walk must continue until it traverses the directed edge closest to vertex 1 which was not traversed during the vertex cover portion of the walk. The expected vertex cover time for such a walk is the same as the amount of time necessary for a walk started at vertex 1 to first arrive at vertex n, i.e. the hitting time from 1 to n. As was mentioned in Chapter 1, $E_1 H_n = (n-1)^2$. It remains then to determine the additional number of steps needed to complete the edge cover. As we contemplated how to go about doing this many questions arose. How many edges, on average, were left uncovered during the vertex cover walk? How far back onto the path was the first missed edge? On average how far back should we expect to go? We answer these questions by first computing the probability, for each edge $(k, k-1)$, that it is the edge closest to vertex 1 not traversed during the vertex cover portion of the walk. Then for each edge we will multiply this probability by the hitting time from vertex n to vertex $k-1$. Finally, adding these quantities yields the expected number of additional steps to complete the edge cover. Once we have this description we will prove that the expected additional number of steps taken after the vertex cover is completed is bounded by a term of lower order than the vertex cover time. Specifically we will establish the following:

Theorem 3.1 *The expected number of steps needed to ensure a simple random walk started at an endpoint of P_n has traversed each edge in the path in both directions is*

$$(1 + o(1))n^2.$$

Indeed, we will actually show that after completing a vertex cover, the expected number of additional steps needed to complete an edge cover is $\theta(n^2/\log n)$.

3.2 The Probability Distribution

For each integer k, $2 \leq k \leq n$, we wish to compute $P(X_k^n)$, the probability that k is the least integer such that edge $(k, k-1)$ was not traversed during the vertex cover portion of the random walk. We will do this by relating the probability distribution for P_n to the probability distribution of P_{n-1}.

Lemma 3.2 *For* $2 \leq k \leq n-1$,

$$P(X_k^n) = \left(\frac{n-k}{n-k+1}\right) P(X_k^{n-1}), \tag{3.1}$$

and

$$P(X_n^n) = \sum_{k=2}^{n-1} \left(\frac{1}{n-k+1}\right) P(X_k^{n-1}). \tag{3.2}$$

The proof of this lemma, and others to follow, will make frequent use of the following easily established result.

Proposition 3.3 *Given a random walk on* P_n *started at vertex* j*, the probability that the walk reaches vertex* 1 *prior to vertex* n *is*

$$\left(\frac{n-j}{n-1}\right).$$

Proof

Let r_j be the probability that starting at vertex j the walk reaches vertex 1 prior to vertex n. Then $r_1 = 1, r_n = 0$ and

$$r_j = \left(\frac{1}{2}\right) r_{j-1} + \left(\frac{1}{2}\right) r_{j+1}.$$

The solution to this recurrence relation is $(n-j)/(n-1)$, as claimed. ■

Proof of Lemma 3.2

Consider a random walk started at vertex 1 on P_n. Upon first visiting vertex $n-1$ this walk completes a vertex cover walk for P_{n-1}. Thus the probability distribution of untraversed edges at this point is $P(X_k^{n-1})$, $2 \leq k \leq n-1$. With probability $(n-k)/(n-k+1)$ the walk will reach vertex n prior to returning to vertex $k-1$. In this case, if k is the least integer such that edge $(k, k-1)$ was not traversed in the walk prior to reaching vertex $n-1$, then it remains so when vertex n is reached. All of these walks then contribute to $P(X_k^n)$. This means that for $2 \leq k \leq n-1$,

$$P(X_k^n) = \left(\frac{n-k}{n-k+1}\right) P(X_k^{n-1}).$$

This leaves only those walks that, upon first reaching vertex $n-1$, step back to vertex $k-1$ prior to their first visit to vertex n. The probability that this occurs is $1/(n-k+1)$. Each of these walks has the property that edge $(n, n-1)$ is the only edge not traversed when the walk first visits vertex n. This gives us

$$P(X_n^n) = \sum_{k=2}^{n-1} \left(\frac{1}{n-k+1}\right) P(X_k^{n-1}). \qquad ■$$

Clearly $P(X_k^n) = 0$ for $n < k$. We may use this fact to reduce $P(X_k^n)$ to a multiple of $P(X_k^k)$. Iterating (3.1) yields

$$P(X_k^n) = \prod_{j=k+1}^{n} \left(\frac{j-k}{j-k+1}\right) P(X_k^k) = \left(\frac{1}{n-k+1}\right) P(X_k^k). \tag{3.3}$$

Using this fact, (3.2) becomes

$$P(X_n^n) = \sum_{k=2}^{n-1} \left(\frac{1}{n-k+1}\right) \left(\frac{1}{n-k}\right) P(X_k^k).$$

Define the sequence $\{f_n\}$ by $f_{n-1} = P(X_n^n)$. Then by equation(3.3)

$$P(X_k^n) = \left(\frac{1}{n-k+1}\right) f_{k-1}. \tag{3.4}$$

But the values of $P(X_k^n)$, for $2 \leq k \leq n$, represent a probability distribution on the edges of P_n, so

$$\sum_{k=2}^{n} P(X_k^n) = 1.$$

This leads to one of the many interesting identities satisfied by the sequence $\{f_n\}$, namely

$$\sum_{j=1}^{n-1} \left(\frac{1}{n-j}\right) f_j = 1. \tag{3.5}$$

This sequence will play a central role in some of the arguments which follow, and we will rely on this identity throughout.

Remark: For future use we will define $f_0 = 0$. A more detailed study of the properties of the sequence $\{f_n\}$ is contained in Appendix A.

It is worth noting our interpretation of the elements of the sequence $\{f_n\}$. When the walk first visits vertex k the probability that it reaches vertex n prior to returning to vertex $k-1$ is $1/(n-k+1)$. Thus looking back to equation(3.4) we see that f_{k-1} represents the conditional probability that, given that the edge $(k, k-1)$ was not traversed prior to the walk reaching vertex n, it is the closest such edge to vertex 1.

Remark: Clearly upon first completing a vertex cover of the path the walk has failed to traverse at least one edge, namely $(n, n-1)$. Based on the observations above we know that the *expected number of edges* which were not traversed during the vertex cover portion of the walk is

$$\sum_{k=2}^{n} \frac{1}{n-k+1} = h_{n-1}.$$

Remark: We may also now determine on average how far back onto the path we must go to complete the edge cover.

$$\sum_{k=2}^{n} (n-k+1) \left(\frac{1}{n-k+1} \right) P(X_k^n) = \sum_{j=1}^{n-1} f_j. \tag{3.6}$$

In Appendix A we prove that $f_n \sim (1/h_n)$, and by combining a lower bound on f_n with techniques similar to those used in Lemma 3.4 prove that (3.6) is $\theta(n/\log n)$. If one could establish this bound on (3.6) directly then Lemma 3.4 would be an immediate consequence.

3.3 Computing the Expected Directed Edge Cover Time

Given the probability distribution discussed in Section 3.2 we can now bound the expected directed edge cover time for P_n. Using the hitting time result of Section 1.2 we find that the hitting time from vertex n to vertex $k-1$ in P_n is $(n-k+1)^2$. If we let E_1C^e denote the expected directed edge cover time for a random walk started at vertex 1 on P_n, then

$$E_1C^e = (n-1)^2 + \sum_{k=2}^{n} (n-k+1)^2 P(X_k^n).$$

Combining this with equation(3.3) yields

$$\begin{aligned} E_1C^e &= (n-1)^2 + \sum_{k=2}^{n} (n-k+1) P(X_k^k) \\ &= (n-1)^2 + \sum_{k=1}^{n-1} (n-k) f_k. \end{aligned} \tag{3.7}$$

So to prove that the expected directed edge cover time is $(1+o(1))n^2$ we need only show that $\sum_{k=1}^{n-1} (n-k) f_k$ is $o(n^2)$.

In order to bound this expected number of additional steps needed after the vertex cover is complete we will use the theory of generating functions. Let the sequence $\{g_n\}$ be defined by

$$g_{n-1} = \sum_{k=1}^{n-1} (n-k) f_k. \tag{3.8}$$

Lemma 3.4 *Given g_{n-1} as in equation(3.8), then $g_{n-1} = O\left(\frac{n^2}{\log n}\right)$.*

Proof

The ordinary generating function for the sequence $\{f_n\}$, $F(x)$, may be found from the identity in equation(3.5). If

$$F(x) = \sum_{k \geq 0} f_k x^k,$$

then from (3.5) we have that

$$F(x) \ln\left(\frac{1}{1-x}\right) = \frac{x^2}{1-x},$$

or

$$F(x) = \frac{-x^2}{(1-x)\ln(1-x)}.$$

Let

$$G(x) = \sum_{n \geq 1} g_n x^n.$$

Then using definition(3.8) we have

$$G(x) = F(x)\frac{1}{(1-x)^2} = \frac{x}{(1-x)^3}\frac{-x}{\ln(1-x)} = \frac{x}{(1-x)^3}B(-x), \tag{3.9}$$

where $B(x) = \sum_{k \geq 0} b_k x^k$ is the ordinary generating function for the Bernoulli numbers of the second kind, $\{b_n\}$, defined and studied by Jordan[17].[1] The

[1] D. Fielder[15] communicated our interest in this generating function to P. Bruckman[8] who pointed out its relation to similar functions in Howard[16], which in turn led us to the work of Jordan.

elements of the sequence $\{b_n\}$ are coefficients in the Bernoulli Polynomials of the 2^{nd} kind given by

$$\Psi_n(x) = \int \binom{x}{n-1} dx = \sum_{k=0}^{n} b_k \binom{x}{n-k}.$$

Equation(3.9) may be written as

$$G(x) = \frac{x}{(1-x)^3} \left(\sum_{k=0}^{\infty} (-1)^k b_k x^k \right),$$

which makes it clear that

$$[x^n]G(x) = \sum_{j=0}^{n-1} \binom{3+j-1}{j} (-1)^{n-j-1} b_{n-j-1}. \tag{3.10}$$

It also follows immediately from (3.9) that

$$\left(\frac{1-x}{x} \right) F(x) = B(-x)$$

which shows that

$$f_{n+1} - f_n = (-1)^n b_n. \tag{3.11}$$

Since $\Delta \Psi_n(x) = \Psi_{n-1}(x)$, where Δ is the difference operator, it follows that the coefficient b_i in $\Psi_n(x)$ is independent of n, and it is easily seen that

$$\int_0^1 \binom{t}{n} dt = \Psi_{n+1}(1) - \Psi_{n+1}(0) = b_n. \tag{3.12}$$

Combining this with (3.11) and $\binom{t}{n} = \binom{t-1}{n} + \binom{t-1}{n-1}$ leads to

$$f_n = (-1)^{n-1} \int_0^1 \binom{t-1}{n-1} dt. \tag{3.13}$$

It also follows from the integral representation of b_n in (3.12) that (3.10) can be written as

$$\begin{aligned} g_n &= \sum_{j=0}^{n-1} \binom{3+j-1}{j} (-1)^{n-j-1} \int_0^1 \binom{t}{n-j-1} dt \\ &= \int_0^1 \sum_{j=0}^{n-1} (-1)^{n-j-1} \binom{3+j-1}{j} \binom{t}{n-j-1} dt. \end{aligned} \tag{3.14}$$

By considering the coefficient of x^n in $\frac{x}{(1-x)^3}(1-x)^t$ we see that

$$\sum_{j=0}^{n-1} (-1)^{n-j-1} \binom{3+j-1}{j} \binom{t}{n-j-1} = (-1)^{n-1} \binom{t-3}{n-1}.$$

So equation(3.14) becomes

$$g_n = \int_0^1 (-1)^{n-1} \binom{t-3}{n-1} dt. \tag{3.15}$$

In order to obtain an upper bound for g_{n-1} note that

$$\begin{aligned} g_{n-1} &= \int_0^1 (-1)^{n-2} \binom{t-3}{n-2} dt \\ &= (-1)^{n-2} \int_0^1 \left(\frac{1}{2}\right) \left(\frac{t-3}{3}\right) \left(\frac{t-4}{4}\right) \cdots \\ &\quad \cdots \left(\frac{t-(n-2)}{n-2}\right) (t-(n-1))(t-n) dt \\ &= \left(\frac{1}{2}\right) \int_0^1 \left(1-\frac{t}{3}\right) \left(1-\frac{t}{4}\right) \cdots \\ &\quad \cdots \left(1-\frac{t}{n-2}\right) (t^2-(2n-1)t+n(n-1)) dt. \end{aligned} \tag{3.16}$$

Using the fact that $(1-x) \leq e^{-x}$, we can bound (3.16) from above by

$$\left(\frac{1}{2}\right) \int_0^1 e^{-t(\frac{1}{3}+\frac{1}{4}+\cdots+\frac{1}{n-2})} (t^2-(2n-1)t+n(n-1)) dt.$$

Then if we let $s = h_{n-2} - \frac{3}{2}$ we have

$$g_{n-1} \leq \left(\frac{1}{2}\right) \int_0^1 e^{-ts}(t^2 - (2n-1)t + n(n-1))dt. \qquad (3.17)$$

Integration by parts in inequality(3.17) leads to

$$g_{n-1} \leq \frac{n^2}{2(h_{n-2} - \frac{3}{2})} \left(1 - \frac{1}{e^{(h_{n-2} - \frac{3}{2})}}\right) + o(1).$$

Hence

$$g_{n-1} \leq \left(\frac{n^2}{2\ln(n-2) - 3}\right) + o(1) = O\left(\frac{n^2}{\ln n}\right). \qquad \blacksquare$$

Proof of Theorem 3.1

The proof of Theorem 3.1 is immediate when one combines equations(3.7) and (3.8) with the result of Lemma 3.4. ■

Remark: In Appendix A we show that $g_{n-1} = \Omega(n^2/\log n)$. So the expected additional number of steps needed to complete a vertex cover walk to an edge cover walk is actually $\theta(n^2/\log n)$.

3.4 Starting from an Arbitrary Point

In Sections 3.1-3.3 we described how one could compute the edge cover time of a simple random walk on P_n started at vertex 1, using coefficients we chose to label f_n. We have that after completing a vertex cover the expected number of additional steps needed to ensure an edge cover is $\theta(n^2/\log n)$. However, the definition of edge cover time is the maximum of the edge cover times calculated for all possible starting points. As was mentioned in Section 1.2, the vertex cover time for a path is maximized for walks started from the center. This was easily seen by maximizing the solution to the recurrence formula mentioned there. It has been conjectured, although still not proven, that the vertex cover time for all trees is maximized by a random walk on the path started at the center. Unfortunately, there does not appear be such a simple description of the edge cover time for a walk on the path. Nonetheless, we have been able to establish a tight bound on the edge cover time for walks started at an endpoint. Now we wish to generalize that result to the case in which we start the walk from an arbitrary vertex $1 \leq j \leq (n+1)/2$. With this information we will answer some of the questions posed above for the edge cover case. Note that due to the symmetry in the path we need not concern ourselves with walks started beyond the center/centers of the graph.

We will prove that for $n \geq 9$ the expected edge cover time on P_n is a strictly increasing function of the starting point, j, for $2 \leq j \leq (n+1)/2$. Combining this with easily computed results for paths on nine or less vertices we will then be able to deduce that the expected edge cover time on a path is maximized in walks started from the center/centers when $n \geq 8$ and from the endpoint for $n < 8$. Moreover we will be able to prove that the additional number of steps needed to complete an edge cover for a walk started at the center/centers of a path is $O(n^2/\log n)$. An interesting fact that we will establish along the way is that the edge cover time for a walk started at an endpoint on $P_n, n \geq 3$, always exceeds that of a walk started at its point of attachment, thus this time is not then minimized by starting at an endpoint. It turns out that this is the case in all graphs, for all leaves, which we will prove in Chapter 4.

Recall that when considering a walk started at vertex 1, our strategy was first to make use of the expected vertex cover time which was nothing more than the hitting time from vertex 1 to vertex n. We then computed for each k, $2 \leq k \leq n$, the probability that k was the least integer such that edge $(k, k-1)$ was not traversed during the vertex cover portion of the walk. Multiplying the probability for each k by the hitting time from n to $k-1$ and adding the results together then gave us the expected number of additional steps necessary to complete an edge cover walk.

Now we need the vertex cover time from an arbitrary starting point.

Proposition 3.5 *On the path, P_n, labeled in the natural way, the vertex cover time for a random walk started at vertex j is*

$$E_j C^v = \frac{5}{4}(n-1)^2 - \frac{1}{4}(n-2j+1)^2. \tag{3.18}$$

As was mentioned in Section 1.2, the proof of this result is similar to the recursive proof that the hitting time from one endpoint to the other is $(n-1)^2$.

Unfortunately no such simple recursive argument appears to be available for the computation of the expected edge cover time. In order to use our previously successful method we must condition the probability distribution on the last endpoint visited during the vertex cover portion of the walk. Once we have found such a conditional probability distribution we can again multiply the probability for each edge by the hitting time from the given endpoint to the far end of that edge. Summing these products will then give us the expected number of additional steps needed.

We will use the following notation.

$X^n_{j,k}$ - the event that k is the least integer such that a random walk on P_n, started at j, upon completing a vertex cover, has not traversed edge $(k, k-1)$.

$Y^n_{j,k}$ - the event that k is the greatest integer such that a random walk on P_n, started at j, upon completing a vertex cover, has not traversed edge $(k, k+1)$.

$A_j^{1\to n}$ - the event that a random walk started at j completes a vertex cover at n, i.e. visits vertex 1 prior to n.

$A_j^{n\to 1}$ - the event that a random walk started at j completes a vertex cover at 1, i.e. visits vertex n prior to 1.

The following statements are immediate.

$$\begin{aligned}
P(Y_{j,k}^n \mid A_j^{1\to n}) &= 0 \qquad \forall k \\
P(X_{j,k}^n \mid A_j^{n\to 1}) &= 0 \qquad \forall k \\
P(X_{j,k}^n \mid A_j^{1\to n}) &= 0 \qquad 1 \le k \le j \\
P(Y_{j,k}^n \mid A_j^{n\to 1}) &= 0 \qquad j \le k \le n-1.
\end{aligned}$$

Then we must determine

$$\begin{aligned}
&P(X_{j,k}^n \mid A_j^{1\to n}) \qquad j+1 \le k \le n \\
&P(Y_{j,k}^n \mid A_j^{n\to 1}) \qquad 1 \le k \le j-1.
\end{aligned}$$

Noting the symmetry of a random walk on P_n we have

$$P(Y_{j,k}^n \mid A_j^{n\to 1}) = P(X_{n-j+1,n-k+1}^n \mid A_{n-j+1}^{1\to n}). \tag{3.19}$$

Thus it will suffice to obtain a characterization of $P(X_{j,k}^n \mid A_j^{1\to n})$.

Lemma 3.6 *For a random walk on P_n, $n \geq 3$, any starting vertex j, and k satisfying $j+1 \leq k \leq n-1$, we have*

$$P(X_{j,k}^{n} \mid A_j^{1\to n}) = \left(\frac{n-j-1}{n-2}\right)\left(\frac{n-1}{n-j}\right)\left(\frac{n-k}{n-k+1}\right) P(X_{j,k}^{n-1} \mid A_j^{1\to n-1})$$

and for $k = n$

$$\begin{aligned} P(X_{j,n}^{n} \mid A_j^{1\to n}) &= \frac{j-1}{(n-2)(n-j)} + \\ &\left(\frac{n-j-1}{n-2}\right)\left(\frac{n-1}{n-j}\right)\sum_{k=j+1}^{n-1} \frac{1}{n-k+1} P(X_{j,k}^{n-1} \mid A_j^{1\to n-1}). \end{aligned}$$

Proof

We are now concentrating on random walks on P_n that start at j and complete a vertex cover upon first stepping to vertex n. Notice that all of these walks complete a vertex cover of P_{n-1} before yielding a vertex cover of P_n. We will use this fact to relate the probability distribution of $P(X_{j,k}^{n-1} \mid A_j^{1\to n-1})$, $j+1 \leq k \leq n-1$, to that of $P(X_{j,k}^{n} \mid A_j^{1\to n})$, $j+1 \leq k \leq n$.

Upon first arriving at n the walk has covered all vertices in P_n and the last new vertex visited prior to n must have been either 1 or $n-1$. We will compute the probability distribution of $P(X_{j,k}^{n} \mid A_j^{1\to n})$, $j+1 \leq k \leq n$, by considering these two subcases separately.

Case(1): Suppose the last new vertex visited prior to n was $n-1$. Then we need to compute the probability that this walk visited 1 prior to $n-1$, given that it visits 1 prior to n. Using elementary probability we have

$$P(A_j^{1\to n-1} \mid A_j^{1\to n}) = \frac{P(A_j^{1\to n-1} \cap A_j^{1\to n})}{P(A_j^{1\to n})}.$$

However, $P(A_j^{1\to n} \mid A_j^{1\to n-1}) = 1$, so for $1 \le j \le n-1$,

$$\begin{aligned} P(A_j^{1\to n-1} \mid A_j^{1\to n}) &= \frac{P(A_j^{1\to n-1})}{P(A_j^{1\to n})} \\ &= \frac{\left(\frac{n-j-1}{n-2}\right)}{\left(\frac{n-j}{n-1}\right)} = \left(\frac{n-j-1}{n-2}\right)\left(\frac{n-1}{n-j}\right). \end{aligned} \tag{3.20}$$

All of these walks that visit 1 prior to $n-1$(and prior to n) begin as vertex cover walks, started at j, for P_{n-1}. So $P(X_{j,k}^{n-1} \mid A_j^{1\to n-1})$ is well defined in these cases. Using Proposition 3.3 we know that upon first reaching vertex $n-1$ the fraction $(n-k)/(n-k+1)$ of them will reach vertex n before returning to vertex $k-1$. Thus we have for $j+1 \le k \le n-1$,

$$P(X_{j,k}^{n} \mid A_j^{1\to n} \cap A_j^{1\to n-1}) = \left(\frac{n-k}{n-k+1}\right) P(X_{j,k}^{n-1} \mid A_j^{1\to n-1}). \tag{3.21}$$

The quantity $P(X_{j,k}^{n} \mid A_j^{1\to n})$ for $j+1 \le k \le n-1$, can now be obtained by making use of

$$P(X_{j,k}^{n} \mid A_j^{1\to n}) = \frac{P(X_{j,k}^{n} \cap A_j^{1\to n} \cap A_j^{1\to n-1}) + P(X_{j,k}^{n} \cap A_j^{1\to n} \cap A_j^{n-1\to 1})}{P(A_j^{1\to n})}.$$

Note that $P(X_{j,k}^{n} \cap A_j^{1\to n} \cap A_j^{n-1\to 1}) = 0$, for $j+1 \le k \le n-1$, from which it follows that

$$P(X_{j,k}^{n} \mid A_j^{1\to n}) = \frac{P(X_{j,k}^{n} \cap A_j^{1\to n} \cap A_j^{1\to n-1})}{P(A_j^{1\to n})}$$

$$= \frac{P(X_{j,k}^{n} \cap A_j^{1\to n} \cap A_j^{1\to n-1})}{P(A_j^{1\to n} \cap A_j^{1\to n-1})} \frac{P(A_j^{1\to n} \cap A_j^{1\to n-1})}{P(A_j^{1\to n})}$$

$$= P(X_{j,k}^{n} \mid A_j^{1\to n} \cap A_j^{1\to n-1}) P(A_j^{1\to n-1} \mid A_j^{1\to n}).$$

Combining this with (3.20) and (3.21) yields for $j+1 \leq k \leq n-1$

$$P(X_{j,k}^{n} \mid A_j^{1\to n}) = \left(\frac{n-j-1}{n-2}\right)\left(\frac{n-1}{n-j}\right)\left(\frac{n-k}{n-k+1}\right) P(X_{j,k}^{n-1} \mid A_j^{1\to n-1}).$$

Within Case (1) this leaves only those walks which reach vertex $k-1$ before vertex n. The probability that this occurs, given that the walk visited 1 prior to $n-1$, is $1/(n-k+1)$. These will contribute to $P(X_{j,n}^{n} \mid A_j^{1\to n})$.

Case (2): Suppose the last new vertex visited prior to n was 1. In this case the walk visited $n-1$, 1, and n, in that order, so we have

$$\begin{aligned} P(A_j^{n-1\to 1} \mid A_j^{1\to n}) &= \frac{P(A_j^{n-1\to 1} \cap A_j^{1\to n})}{P(A_j^{1\to n})} \\ &= \frac{\left(\frac{j-1}{n-2}\right)\left(\frac{1}{n-1}\right)}{\left(\frac{n-j}{n-1}\right)} \\ &= \frac{j-1}{(n-2)(n-j)}. \end{aligned} \tag{3.22}$$

Notice that the only edge which these walks have not traversed upon reaching n is $(n, n-1)$. Thus all of these walks belong to the class of walks that complete a vertex cover at n and have $k = n$ as the least integer such that edge $(k, k-1)$ has not been traversed.

Combining this fraction with the walks described in the latter part of Case (1) results in

$$\begin{aligned} P(X_{j,n}^{n} \mid A_j^{1\to n}) &= P(A_j^{n-1\to 1} \mid A_j^{1\to n}) + \\ &\sum_{k=j+1}^{n-1} \frac{1}{n-k+1} P(X_{j,k}^{n-1} \mid A_j^{1\to n-1}) \left(\frac{n-j-1}{n-2}\right)\left(\frac{n-1}{n-j}\right). \end{aligned}$$

Using equation(3.22) we have

$$\begin{aligned} &P(X_{j,n}^{n} \mid A_j^{1\to n}) = \\ &\frac{j-1}{(n-2)(n-j)} + \\ &\left(\frac{n-j-1}{n-2}\right)\left(\frac{n-1}{n-j}\right) \sum_{k=j+1}^{n-1} \left(\frac{1}{n-k+1}\right) P(X_{j,k}^{n-1} \mid A_j^{1\to n-1}). \quad \blacksquare \end{aligned}$$

3.5 Relation between $P(X_{j,k}^{k} \mid A_j^{1\to k})$ and $\{f_n\}$

Lemma(3.6) provides us with recursive formulas for the probabilities we need in terms of the probability distribution on a path containing 1 less vertex. This is not very useful in practice. Notice that if we set $j = 1$ we have exactly the recurrence that we found in our analysis of walks started from vertex 1.

This suggests that we may be able to reduce our recursive expression to a more tractable form. Rearranging and iterating we have for $j+1 \leq k \leq n-1$,

$$\begin{aligned}
P(X^n_{j,k} \mid A^{1\to n}_j) &= \\
&= \left(\frac{n-k}{n-k+1}\right)\left(\frac{n-j-1}{n-j}\right)\left(\frac{n-1}{n-2}\right) P(X^{n-1}_{j,k} \mid A^{1\to n-1}_j) \\
&= \left(\frac{n-1}{n-j}\right)\left(\frac{1}{n-k+1}\right)(n-k-1)\left(\frac{n-j-2}{n-3}\right) P(X^{n-2}_{j,k} \mid A^{1\to n-2}_j) \\
&= \quad \vdots \\
&= \left(\frac{n-1}{n-j}\right)\left(\frac{1}{n-k+1}\right)\left(\frac{k-j}{k-1}\right) P(X^{k}_{j,k} \mid A^{1\to k}_j) \\
&= \left(\frac{n-1}{n-j}\right)\left(\frac{1}{n-k+1}\right) P(X^{k}_{j,k} \mid A^{1\to k}_j) P(A^{1\to k}_j). \qquad (3.23)
\end{aligned}$$

$P(X^n_{j,k} \mid A^{1\to n}_j) = 0$ for $n \leq k$ so this is as far as we may reduce. Also, since the $P(X^n_{j,k} \mid A^{1\to n}_j)$, for $j+1 \leq k \leq n$, represent a probability distribution we have

$$\begin{aligned}
P(X^n_{j,n} \mid A^{1\to n}_j) & \\
&= 1 - \sum_{k=j+1}^{n-1} P(X^n_{j,k} \mid A^{1\to n}_j) \\
&= 1 - \left(\frac{n-1}{n-j}\right)\sum_{k=j+1}^{n-1}\left(\frac{1}{n-k+1}\right) P(X^{k}_{j,k} \mid A^{1\to k}_j) P(A^{1\to k}_j). \qquad (3.24)
\end{aligned}$$

These results have reduced the amount of information we need in order to use the stated probabilities to determine the expected edge cover time starting from j. Our next step is to find a description of $P(X^k_{j,k} \mid A^{1\to k}_j)$ in terms of quantities we already know.

In Section 3.3 we provided a tight bound on the expected number of additional steps needed to complete an edge cover having already accomplished a vertex cover starting from an endpoint. Our bound relied on a connection between the probability that a particular edge was the first missed during the vertex cover walk and our coefficients f_n. It seems that there should be some connection between the edge probabilities we have calculated for the more general case and the f_n.

Remember that

$$P(A_j^{1\to n}) = \binom{n-j}{n-1}.$$

Combining this with equation(3.24) leads to the equation

$$\begin{aligned} P(X_{j,n}^n \mid A_j^{1\to n})P(A_j^{1\to n}) &= \binom{n-j}{n-1} \\ &- \sum_{k=j+1}^{n-1} \left(\frac{1}{n-k+1}\right) P(X_{j,k}^k \mid A_j^{1\to k})P(A_j^{1\to k}) \end{aligned}$$

or

$$\binom{n-j}{n-1} = \sum_{k=j+1}^{n} \left(\frac{1}{n-k+1}\right) P(X_{j,k}^k \mid A_j^{1\to k})P(A_j^{1\to k}). \tag{3.25}$$

When $j = 1$ this is the recurrence which helped define the coefficients f_n in equation(3.5).

Lemma 3.7 *For fixed k,*

$$P(X_{j,k}^k \mid A_j^{1\to k})P(A_j^{1\to k}) = \sum_{r=1}^{k-j} (f_r - f_{r-1}) \binom{k-j-r+1}{k-r}. \tag{3.26}$$

Proof

We will prove the lemma by showing that, for fixed k, the two sequences dependent on j are the same for $j = 1$ and satisfy the same defining recurrence relation, given in equation(3.25).

Consider the case when $j = 1$. Then

$$P(X_{1,k}^{k} \mid A_1^{1\to k})P(A_1^{1\to k}) = P(X_{1,k}^{k}) = f_{k-1}$$

where the first equality is due to the fact that $P(A_1^{1\to k}) = 1$, and the second comes from the definition of f_{k-1}. But, remembering that f_0 is defined to be 0,

$$\sum_{r=1}^{k-1} (f_r - f_{r-1}) \left(\frac{k-r}{k-r}\right) = f_{k-1}.$$

So the claim is true when $j = 1$.

Putting $s = k - j$ we have

$$\begin{aligned}
&\sum_{k=j+1}^{n} \left(\frac{1}{n-k+1}\right) \sum_{r=1}^{k-j} (f_r - f_{r-1}) \left(\frac{k-j-r+1}{k-r}\right) \\
&= \sum_{s=1}^{n-j} \left(\frac{1}{n-s-j+1}\right) \left[\sum_{r=1}^{s} (f_r - f_{r-1}) \left(\frac{s-r+1}{s-r+j}\right)\right], \\
&= \sum_{r=1}^{n-j} \left(\frac{n-j-r+1}{n-r}\right) \left[\sum_{s=1}^{r} (f_s - f_{s-1}) \left(\frac{1}{r-s+1}\right)\right]. \qquad (3.27)
\end{aligned}$$

For $r \geq 1$

$$\begin{aligned}
\sum_{s=1}^{r} (f_s - f_{s-1}) \left(\frac{1}{r-s+1}\right) &= \sum_{s=1}^{r} f_s \left(\frac{1}{(r+1)-s}\right) \\
&- \sum_{s=1}^{r} f_{s-1} \left(\frac{1}{r-(s-1)}\right),
\end{aligned}$$

and relying on the definition of $\{f_n\}$, we know since $f_0 = 0$

$$\sum_{s=1}^{r} f_{s-1}\left(\frac{1}{r-(s-1)}\right) = \sum_{x=1}^{r-1} f_x\left(\frac{1}{r-x}\right) = 1.$$

So for $r > 1$ the inner sum in (3.27) is identically 0 leaving only the term $r = 1$ which is $(n-j)/(n-1)$. ■

3.6 Analysis of the Edge Cover Time for an Arbitrary Start Vertex

We now have the probability that a particular edge is the last edge traversed in an edge cover walk on P_n started at vertex j expressed strictly in terms of weighted sums of coefficients we discovered in our discussion of such a walk started at vertex 1. Our next task is to use this information to prove that the expected edge cover time on a path started at vertex j is an increasing function of j for $2 \leq j \leq (n+1)/2$. Once we have shown this, then by the symmetry of P_n, we will have that the expected edge cover time for P_n is maximized when a walk is started at the center, or when n is even, centers, of the path. While studying this question we combined the ability of *Mathematica* to do exact rational number arithmetic, and the probability distribution defined in Lemma 3.7 based on weighted sums of the f_n, to confirm computationally that the expected edge cover time on P_n, started

					Start Vertex			
n	1	2	3	4	5	25	50	75
2	2							
3	6.5	6						
4	13.42	12.75						
5	22.71	21.96	22					
6	34.35	33.55	33.76					
7	48.32	47.49	47.92	48.16				
8	64.60	63.75	64.43	64.97				
9	83.19	82.32	83.25	84.14	84.48			
10	104.08	103.19	104.37	105.63	106.36			
49	2668.67	2667.69	2677.66	2692.03	2708.61	2939.39		
149	24507.3	24506.3	24534.5	24576.1	24626	25977	27297	27763

Table 3.1: Values of Edge Cover Time

at vertex j, is an increasing function of j for $2 \leq j \leq (n+1)/2$ when $5 \leq n \leq 150$. A sample of the actual predicted edge cover times is contained in Table 3.1.

Let E_jC^e be the expected edge cover time for a walk on P_n started at vertex j. Then E_jC^e may be computed by finding the expected vertex cover time for a walk started at j, E_jC^v, and then computing the expected number of additional steps the walk will need to return from the last point of the vertex cover to the edge closest to the starting point which was not traversed during the vertex cover portion of the walk. In the notation previously introduced this is then

$$
\begin{aligned}
E_jC^e = E_jC^v \quad &+ \sum_{k=j+1}^{n} (n-k+1)^2\, P(X^n_{j,k} \mid A_j^{1\to n})P(A_j^{1\to n}) \\
&+ \sum_{k=1}^{j-1} (k)^2\, P(Y^n_{j,k} \mid A_j^{n\to 1})P(A_j^{n\to 1}).
\end{aligned}
$$

Using equation(3.19) this becomes

$$
\begin{aligned}
E_jC^e = E_jC^v \quad &+ \sum_{k=j+1}^{n} (n-k+1)^2\, P(X^n_{j,k} \mid A_j^{1\to n})P(A_j^{1\to n}) \\
&+ \sum_{k=1}^{j-1} (k)^2\, P(X^n_{n-j+1,n-k+1} \mid A_{n-j+1}^{1\to n})P(A_{n-j+1}^{1\to n}).
\end{aligned}
$$

Combining this with equations(3.23) and (3.24), and making use of

$$
E_jC^v = \left(\frac{5}{4}\right)(n-1)^2 - \left(j - \left(\frac{n+1}{2}\right)\right)^2
$$

and

$$
P(A_j^{1\to n}) = \left(\frac{n-j}{n-1}\right)
$$

leads to

$$
\begin{aligned}
E_jC^e \quad &= \quad \left(\frac{5}{4}\right)(n-1)^2 - \left(j - \frac{n+1}{2}\right)^2 \\
&+ \sum_{k=j+1}^{n} (n-k+1)\, P(X^k_{j,k} \mid A_j^{1\to k})P(A_j^{1\to k}) \\
&+ \sum_{k=1}^{j-1} (k)\, P(X^{n-k+1}_{n-j+1,n-k+1} \mid A_{n-j+1}^{1\to n-k+1})P(A_{n-j+1}^{1\to n-k+1}).
\end{aligned}
$$

Now using Lemma 3.7 we have

$$
E_jC^e \quad = \quad \left(\frac{5}{4}\right)(n-1)^2 - \left(j - \frac{n+1}{2}\right)^2 +
$$

$$+\sum_{k=j+1}^{n}(n-k+1)\left[-\sum_{r=1}^{k-j}(f_{r-1}-f_r)\left(\frac{k-j-r+1}{k-r}\right)\right]$$
$$+\sum_{k=1}^{j-1}(k)\left[-\sum_{r=1}^{j-k}(f_{r-1}-f_r)\left(\frac{j-k-r+1}{n-k-r+1}\right)\right]. \tag{3.28}$$

In Chapter 4 we will show that in all graphs the expected edge cover time for a random walk started from a leaf exceeds the expected edge cover time for a walk started from its point of attachment by exactly 1 minus the probability that the last new edge visited in a random walk started from that leaf is the edge from its point of attachment to itself. Based on our interpretation of the coefficients f_n we know that for a random walk on P_n started at vertex 1 the probability we complete an edge cover on edge $(2,1)$ is exactly $1/(n-1)$. Thus a random walk started at vertex 2 will have expected edge cover time $(n-2)/(n-1)$ steps less than a walk started at 1. This differs from the vertex cover case where the expected vertex cover on P_n, considered as a function of the starting vertex j, is a strictly increasing function of j for $1 \leq j \leq (n+1)/2$. The exact difference in the expected edge cover times starting from vertices 1 and 2 is confirmed by a direct calculation of equation(3.28) with $j=1$ and $j=2$. It is also worth noting that using the analysis to follow, the expected edge cover time for a random walk on P_n started at vertex 3 exceeds that for a walk started at vertex 2 when $n \geq 5$, however, it does not exceed that for a walk started at vertex 1 until $n \geq 9$.

Theorem 3.8 *On the path P_n, E_jC^e is a strictly increasing function for $2 \leq j \leq \frac{n+1}{2}$ and $n \geq 5$.*

Proof

In order to prove the theorem it suffices to show that $E_{j+1}C^e - E_jC^e > 0$ for $2 \leq j \leq (n-1)/2$. Also since we have established that the theorem is true computationally for $n \leq 150$ we will only need to show this for $n > 150$. We will do so by studying the form of this difference in light of our knowledge of the sequence f_n. We will derive a sequence of inequalities each of whose proof would ensure our result.

Using equation(3.28), and after a good deal of algebra, one obtains

$$\begin{aligned} E_{j+1}C^e - E_jC^e &= (n-2j) - \left(\frac{n-j}{j}\right) + \left(\frac{j}{n-j}\right) \\ &+ \sum_{k=j+2}^{n} (n-k+1) \left[\sum_{r=1}^{k-j} (f_{r-1} - f_r) \left(\frac{1}{k-r}\right)\right] \\ &+ \sum_{k=1}^{j-1} k \left[\sum_{r=1}^{j-k+1} (f_{r-1} - f_r) \left(\frac{-1}{n-k-r+1}\right)\right]. \quad (3.29) \end{aligned}$$

Reversing the order of summation and reindexing the first of the two sums above yields

$$\begin{aligned} E_{j+1}C^e - E_jC^e &= (n-2j) - \left(\frac{n-j}{j}\right) + \left(\frac{j}{n-j}\right) \\ &+ \sum_{k=1}^{n-j-1} k \left[\sum_{r=1}^{n-j-k+1} (f_{r-1} - f_r) \left(\frac{1}{n-k-r+1}\right)\right] \end{aligned}$$

$$+\sum_{k=1}^{j-1} k\left[\sum_{r=1}^{j-k+1}(f_{r-1}-f_r)\left(\frac{-1}{n-k-r+1}\right)\right].$$

It is now easy to see that combining the like terms in the two summations results in

$$\begin{aligned}E_{j+1}C^e - E_jC^e &= (n-2j)-\left(\frac{n-j}{j}\right)+\left(\frac{j}{n-j}\right)\\ &+\sum_{k=1}^{j-1} k\left[\sum_{r=j-k+2}^{n-j-k+1}(f_{r-1}-f_r)\left(\frac{1}{n-k-r+1}\right)\right]\\ &+\sum_{k=j}^{n-j-1} k\left[\sum_{r=1}^{n-j-k+1}(f_{r-1}-f_r)\left(\frac{1}{n-k-r+1}\right)\right].\end{aligned}$$

Focusing on the first of the two sums above and using the fact that the f_n are a decreasing sequence for $n \geq 1$, it is apparent that

$$\sum_{k=1}^{j-1} k\left[\sum_{r=j-k+2}^{n-j-k+1}(f_{r-1}-f_r)\left(\frac{1}{n-k-r+1}\right)\right] > 0.$$

So to prove that $E_{j+1}C^e - E_jC^e > 0$ it suffices to show that

$$\begin{aligned}&(n-2j)\left(\frac{n-j}{j}\right)+\left(\frac{j}{n-j}\right)\\ &\quad\geq \sum_{k=j}^{n-j-1} k\left[\sum_{r=1}^{n-j-k+1}(f_{r-1}-f_r)\left(\frac{-1}{n-k-r+1}\right)\right].\end{aligned} \tag{3.30}$$

Using identity(3.5) we obtain

$$\sum_{r=1}^{n-j-k+1}(f_{r-1}-f_r)\left(\frac{-1}{n-k-r+1}\right)=\sum_{r=n-j-k+2}^{n-k}(f_{r-1}-f_r)\left(\frac{1}{n-k-r+1}\right).$$

Applying this to the right side of inequality(3.30) yields

$$\sum_{k=j}^{n-j-1}(k)\sum_{r=n-j-k+2}^{n-k}(f_{r-1}-f_r)\left(\frac{1}{n-k-r+1}\right)$$

$$= \sum_{k=j}^{n-j-1} (k) \sum_{i=1}^{j-1} \left(\frac{1}{i}\right) (f_{n-k-i} - f_{n-k-i+1}) \tag{3.31}$$

It should be noted that in this form each of the inner sums have exactly $j-1$ terms and that there are a total of $n-2j$ terms in the outer sum. Reversing the order of summation leads to a good deal of cancellation and so (3.31) becomes:

$$\begin{aligned} j\sum_{i=1}^{j-1} \left(\frac{-1}{i}\right) f_{n-j-i+1} \quad &+ \quad \sum_{i=1}^{j-1} \left(\frac{-1}{i}\right) \sum_{k=j+1}^{n-j-1} (f_{k-i+1}) \\ &+ \quad (n-j-1) \sum_{i=1}^{j-1} \left(\frac{1}{i}\right) f_{j-i+1}. \end{aligned}$$

Using identity(3.5) the third term in this expression for (3.31) can be written as

$$(n-j-1) \sum_{i=1}^{j-1} \left(\frac{1}{i}\right) f_{j-i+1} = (n-j-1)\left(1 - \frac{1}{j}\right).$$

Subtracting this term from both sides of (3.30) and multiplying the resulting inequality by -1 yields

$$j + \frac{1}{j} - \frac{n}{n-j} \ \leq\ j\sum_{i=1}^{j-1} \left(\frac{1}{i}\right) f_{n-j-i+1} + \sum_{i=1}^{j-1} \left(\frac{1}{i}\right) \sum_{k=j+1}^{n-j-1} (f_{k-i+1}). \tag{3.32}$$

Let E denote the first term on the right side of inequality(3.32) and F the second term. Then in order to prove that $E_{j+1}C^e - E_jC^e > 0$ it suffices to prove

$$F \geq j + \frac{1}{j} - \frac{n}{n-j} - E. \tag{3.33}$$

Once again using identity(3.5) we have

$$E = j\left[1 - \sum_{k=1}^{n-2j+1}\left(\frac{1}{n-j-k+1}\right)f_k\right].$$

Note that $f_1 = 1$ so

$$E = j - \frac{j}{n-j} - j\sum_{k=2}^{n-2j+1}\left(\frac{1}{n-j-k+1}\right)f_k,$$

and (3.33) may be rewritten as

$$F \geq \frac{1}{j} - 1 + j\sum_{k=2}^{n-2j+1}\left(\frac{1}{n-j-k+1}\right)f_k. \tag{3.34}$$

In order to prove that this inequality holds we now consider the following two cases:

Case (1): $\frac{n}{4} \leq j \leq \frac{n-1}{2}$. Case (2): $2 \leq j < \frac{n}{4}$.

Case 1

Studying the right side of inequality(3.34) and remembering that $f_r \leq \frac{1}{2}$, $\forall r \geq 2$ it is easy to see that

$$j\sum_{k=2}^{n-2j+1}\left(\frac{1}{n-j-k+1}\right)f_k \leq \sum_{k=2}^{n-2j+1} f_k \leq \frac{n}{2} - j.$$

Thus for Case 1 it suffices to show that

$$F \geq \frac{1}{j} - 1 + \frac{n}{2} - j, \quad \text{when } \frac{n}{4} \leq j \leq \frac{n-1}{2}. \tag{3.35}$$

Changing the order of summation in $F = \sum_{i=1}^{j-1} \left(\frac{1}{i}\right) \sum_{k=j+1}^{n-j-1} (f_{k-i+1})$ yields

$$F = \sum_{k=j+1}^{n-j-1} \sum_{i=1}^{j-1} \left(\frac{1}{i}\right) (f_{k-i+1}) .$$

Remembering that h_n denotes the n^{th} harmonic number, we note that

$$\begin{array}{rcll}
\sum_{i=1}^{j-1} \left(\frac{1}{i}\right) (f_{k-i+1}) & \geq & h_{j-1} f_{j+1} & \text{when } k = j+1, \\
\sum_{i=1}^{j-1} \left(\frac{1}{i}\right) (f_{k-i+1}) & \geq & h_{j-1} f_{j+2} & \text{when } k = j+2, \\
\vdots & & \vdots & \\
\sum_{i=1}^{j-1} \left(\frac{1}{i}\right) (f_{k-i+1}) & \geq & h_{j-1} f_{n-j-1} & \text{when } k = n-j-1.
\end{array}$$

Summing these inequalities yields

$$F \geq h_{j-1}(n - 2j - 1) f_{n-j-1}.$$

In Appendix A we show that for $r \geq 3$, $f_r > \left(\frac{3}{4}\right)\left(\frac{1}{h_r}\right)$. So for $n > 150$ and $\frac{n}{4} \leq j \leq \frac{n-1}{2}$ it is certainly true that $f_{n-j-1} > \left(\frac{3}{4}\right) \frac{1}{h_{n-j-1}}$, which implies

$$F > \left(\frac{3}{4}\right) \left(\frac{h_{j-1}}{h_{n-j-1}}\right) (n - 2j - 1). \tag{3.36}$$

A simple application of the Cauchy-Schwartz inequality to h_{j-1} and h_{n-j-1} shows that

$$\frac{h_{j-1}}{h_{n-j-1}} \geq 1 - \frac{h_{n-j-1} - h_{j-1}}{h_{j-1}}.$$

But $h_{n-j-1} - h_{j-1} < \ln 4$ when $\frac{n}{4} \leq j \leq \frac{n-1}{2}$ and $n > 150$, leading to the conclusion that

$$F > \left(\frac{3}{4}\right)\left(1 - \frac{\ln 4}{h_{j-1}}\right)(n - 2j - 1). \tag{3.37}$$

If we let $j = \frac{n}{2} - \alpha$, where $\frac{1}{2} \leq \alpha \leq \frac{n}{4}$, inequality(3.37) becomes

$$F > \left(\frac{3}{4}\right)\left(1 - \frac{\ln 4}{h_{j-1}}\right)(2\alpha - 1). \tag{3.38}$$

and we have

$$\frac{1}{j} - 1 + \left(\frac{n}{2} - j\right) = \frac{1}{j} - 1 + \alpha.$$

Thus to prove inequality(3.35) it suffices to show that

$$\left(\frac{3}{4}\right)\left(1 - \frac{\ln 4}{h_{j-1}}\right)(2\alpha - 1) \geq \frac{4}{n} - 1 + \alpha, \quad \text{for } \frac{1}{2} \leq \alpha = \frac{n}{2} - j \leq \frac{n}{4}. \tag{3.39}$$

For $n > 150$ and $j \geq \frac{n}{4}$, $\frac{\ln 4}{h_{j-1}} < \frac{1}{3}$. So

$$\left(\frac{3}{4}\right)\left(1 - \frac{\ln 4}{h_{j-1}}\right)(2\alpha - 1) > \alpha - \frac{1}{2},$$

and for $n > 150$

$$\alpha - \frac{1}{2} > \alpha - 1 + \frac{4}{n}.$$

This then implies inequality(3.39) which was sufficient to prove that

$$E_{j+1}C^e - E_jC^e > 0 \qquad \text{for} \qquad \frac{n}{4} \leq j \leq \frac{n-1}{2}.$$

Case 2

Now we wish to show the result for $2 \leq j < \frac{n}{4}$. Simple manipulation of our general description of expected edge cover time in equation(3.28) with $j = 2$ and $j = 3$ shows that for $n \geq 5$, $E_3C^e > E_2C^e$. So we really only need to prove $E_{j+1}C^e - E_jC^e > 0$ for $j \geq 3$. Once again referring to inequality(3.34), note that applying identity(3.5) (and reindexing) we have

$$\begin{aligned}\sum_{k=2}^{n-2j+1} \left(\frac{1}{n-j-k+1}\right) f_k &= 1 - \left(\frac{1}{n-j}\right) f_1 - \sum_{k=1}^{j-1} \left(\frac{1}{k}\right) f_{n-j+1-k} \\ &< 1 - \frac{1}{n-j}.\end{aligned}$$

Thus to establish inequality(3.34) it suffices to show that

$$F \geq \frac{1}{j} - 1 + j - \frac{j}{n-j}.$$

From inequality(3.36) we have that

$$F > \left(\frac{3}{4}\right)\left(\frac{h_{j-1}}{h_{n-j-1}}\right)(n-2j-1),$$

so it suffices to prove that

$$\left(\frac{3}{4}\right)\left(\frac{h_{j-1}}{h_{n-j-1}}\right)(n-2j-1) \geq \frac{1}{j} - 1 + j - \frac{j}{n-j}$$

or even

$$\left(\frac{3}{4}\right)\frac{h_{j-1}}{h_{n-j-1}}(n-2j-1) \geq j. \tag{3.40}$$

Let $j = \alpha(n-1)$ with $\frac{3}{n-1} \leq \alpha \leq \frac{1}{4}$ (these bounds are appropriate since here in case (2) $j < \frac{n}{4}$). Then referring to inequality(3.40) and remembering that $h_n > \ln n$, $\forall n \geq 1$, we want

$$\left(\frac{3}{4}\right)\left(\frac{\ln[\alpha(n-1)-1]}{h_{(1-\alpha)(n-1)}}\right)(1-2\alpha)(n-1) > \alpha(n-1).$$

Canceling $n-1$ from both sides we need

$$\left(\frac{3}{4}\right)\left(\frac{\ln[\alpha(n-1)-1]}{h_{(1-\alpha)(n-1)}}\right)(1-2\alpha) \geq \alpha.$$

But $h_k < \ln(k)+1$ for $k > 1$ so it suffices to show that

$$\left(\frac{3}{4}\right)\left(\frac{\ln[\alpha(n-1)-1]}{\ln[(1-\alpha)(n-1)]+1}\right)(1-2\alpha) \geq \alpha.$$

or equivalently

$$\left(\frac{3}{4}\right)(1-2\alpha)\ln[\alpha(n-1)-1] - \alpha(\ln[(1-\alpha)(n-1)]+1) > 0 \qquad (3.41)$$

When $n > 150$ and $\alpha = \frac{3}{n-1}$ or $\alpha = \frac{1}{4}$ the inequality is valid proving that the expression on the left-hand-side of (3.41) is positive at the endpoints of the interval in question. Using basic calculus one finds that the second derivative of this expression, taken with respect to α, is negative throughout the interval, and thus the expression must be positive throughout the interval. By our chain of sufficient inequalities we then have

$$E_{j+1}C^e - E_jC^e > 0 \qquad \text{for} \qquad 2 \leq j < \frac{n}{4}.$$

Cases (1) and (2) in combination show that for $n > 150$, E_jC^e is a strictly increasing function of j for $2 \leq j \leq \frac{n+1}{2}$. ∎

Combining the fact that $E_3C^e > E_1C^e$ for $n \geq 9$ with our computation of the expected edge cover time for small n we have as an immediate corollary that it is maximized by starting at the endpoint for walks on P_n when $n < 8$, and maximized by starting at the center/centers for all longer paths.

3.7 The Additional Number of Steps Needed When Starting From the Center

Equation(3.28) provides a formula to compute the expected edge cover time starting from any point on the path using the vertex cover time and weighted sums of our coefficients f_n. In Section 3.3 we proved, for a random walk started at an endpoint that the additional number of steps needed to assure an edge cover, after the vertex cover walk was completed, was $\theta(n^2/\log n)$. We can prove that this quantity provides an upper bound for walks started from the center/centers.

Theorem 3.9 *For a random walk started at the center/centers of P_n, the expected additional number of steps which are needed to complete an edge cover after all vertices have been visited at least once is $O(n^2/\log n)$.*

Proof

We will first prove that the theorem is true for n odd. We will then use that fact to make the argument for even n. The quantity in which we are interested is $E_jC^e - E_jC^v$ when $j = (n+1)/2$. This consists of the latter two sums in equation(3.28).

$$\begin{aligned} E_jC^e - E_jC^v &= \sum_{k=j+1}^{n} (n-k+1)\left[-\sum_{r=1}^{k-j} (f_{r-1}-f_r)\left(\frac{k-j-r+1}{k-r}\right)\right] \\ &+ \sum_{k=1}^{j-1} k\left[-\sum_{r=1}^{j-k} (f_{r-1}-f_r)\left(\frac{j-k-r+1}{n-k-r+1}\right)\right]. \end{aligned} \tag{3.42}$$

Reindexing the first summation we have

$$\begin{aligned} &\sum_{k=j+1}^{n} (n-k+1)\left[-\sum_{r=1}^{k-j} (f_{r-1}-f_r)\left(\frac{k-j-r+1}{k-r}\right)\right] \\ &= \sum_{k=1}^{n-j} k\left[-\sum_{r=1}^{(n-j+1)-k} (f_{r-1}-f_r)\left(\frac{n-j-r-k+2}{n-r-k+1}\right)\right]. \end{aligned} \tag{3.43}$$

We are interested in the quantity when $j = (n+1)/2$, which implies $n-j = j-1 = (n-1)/2$. So in this case

$$E_jC^e - E_jC^v = 2\sum_{k=1}^{\frac{n-1}{2}} k\left[-\sum_{r=1}^{\frac{n+1}{2}-k} (f_{r-1}-f_r)\left(\frac{\frac{n+1}{2}-k-r+1}{n-k-r+1}\right)\right]. \tag{3.44}$$

It is easier to analyze equation(3.44) if we break up the summation yielding

$$\begin{aligned} \left(\frac{1}{2}\right) E_jC^e - E_jC^v &= \sum_{k=1}^{\frac{(n-1)}{2}} (-k)\left(\sum_{r=1}^{\frac{n+1}{2}-k} f_{r-1}\left(\frac{\frac{n+1}{2}-k-r+1}{n-k-r+1}\right)\right) \\ &+ \sum_{k=1}^{\frac{(n-1)}{2}} k\left(\sum_{r=1}^{\frac{n+1}{2}-k} f_r\left(\frac{\frac{n+1}{2}-k-r+1}{n-k-r+1}\right)\right) \end{aligned}$$

This in turn may be broken up further into

$$\begin{aligned}
\left(\frac{1}{2}\right) E_jC^e - E_jC^v &= -\sum_{k=1}^{\frac{n-1}{2}} kf_0 \begin{pmatrix} \frac{n+1}{2} - k \\ n-k \end{pmatrix} \\
&- \sum_{k=1}^{\frac{n-1}{2}-1} k \sum_{r=2}^{\frac{n+1}{2}-k} f_{r-1} \begin{pmatrix} \frac{n+1}{2} - k - r + 1 \\ n-k-r+1 \end{pmatrix} \\
&+ \sum_{r=1}^{\frac{n-1}{2}} f_r \begin{pmatrix} \frac{n+1}{2} - r \\ n-r \end{pmatrix} \\
&+ \sum_{k=2}^{\frac{n-1}{2}} k \sum_{r=1}^{\frac{n+1}{2}-k} f_r \begin{pmatrix} \frac{n+1}{2} - k - r + 1 \\ n-k-r+1 \end{pmatrix}.
\end{aligned}$$

Now we use the fact that $f_0 = 0$ and reindex again to obtain

$$\begin{aligned}
&\left(\frac{1}{2}\right) E_jC^e - E_jC^v = \\
&- \sum_{k=1}^{\frac{n-1}{2}-1} k \sum_{r=1}^{\frac{n-1}{2}-k} f_r \begin{pmatrix} \frac{n+1}{2} - k - r \\ n-k-r \end{pmatrix} \\
&+ \sum_{r=1}^{\frac{n-1}{2}} f_r \begin{pmatrix} \frac{n+1}{2} - r \\ n-r \end{pmatrix} + \sum_{k=1}^{\frac{n-1}{2}-1} (k+1) \sum_{r=1}^{\frac{n-1}{2}-k} f_r \begin{pmatrix} \frac{n+1}{2} - k - r \\ n-k-r \end{pmatrix}.
\end{aligned}$$

This then reduces to

$$\begin{aligned}
E_jC^e - E_jC^v &= 2 \sum_{k=0}^{\frac{n-1}{2}-1} \sum_{r=1}^{\frac{n-1}{2}-k} f_r \begin{pmatrix} \frac{n+1}{2} - k - r \\ n-k-r \end{pmatrix} \\
&= \sum_{k=0}^{\frac{n-1}{2}-1} \sum_{r=1}^{\frac{n-1}{2}-k} f_r \begin{pmatrix} n+1-2k-2r \\ n-k-r \end{pmatrix} \\
&= 2 \sum_{k=0}^{\frac{n-1}{2}-1} \sum_{r=1}^{\frac{n-1}{2}-k} f_r \\
&- (n-1) \sum_{k=0}^{\frac{n-1}{2}-1} \sum_{r=1}^{\frac{n-1}{2}-k} f_r \begin{pmatrix} 1 \\ n-k-r \end{pmatrix} \qquad (3.45)
\end{aligned}$$

$$
\begin{aligned}
&= 2\sum_{r=1}^{\frac{n-1}{2}} f_r\left(\frac{n+1}{2}-r\right) \\
&- (n-1)\sum_{r=1}^{\frac{n-1}{2}} f_r\left(h_{n-r}-h_{\frac{n-1}{2}}\right) \qquad (3.46)
\end{aligned}
$$

Referring to Appendix A we know that

$$2\sum_{r=1}^{\frac{n-1}{2}} f_r\left(\frac{n+1}{2}-r\right) = 2g_{\left(\frac{n+1}{2}\right)}.$$

So

$$E_jC^e - E_jC^v \leq 2g_{\left(\frac{n+1}{2}\right)} = O\left(\frac{n^2}{\log n}\right).$$

Now let n be even. Consider a random walk started at vertex $n/2-1$ on P_n and another walk started at $n/2-1$ on P_{n+1}. It becomes clear that the expected edge cover time of the latter must be at least as great as the former when one considers the following modification. Modify P_n by adding a loop to vertex n. Now upon reaching vertex n let the walk have probability $1/2$ of stepping back to vertex $n-1$ and probability $1/2$ of stepping around the loop. Compare a walk started at vertex $n/2-1$ on this modified path to that of the walk on P_n. Each time we visit vertex n on P_n we return to vertex $n-1$(and therefore back towards untraversed edges) with probability 1 taking 1 step to do so. On the modified path we step back to vertex $n-1$ only $1/2$ of the time and the other $1/2$ of the time waste a step on the loop. It is clear that in the modified case the expected number of steps to traverse all the directed edges from 1 to n must be at least as great as the expectation on P_n. Thus it is clear that when one compares the edge cover walks on P_n and P_{n+1}, as described above, that the expected edge cover time for the walk on P_n must bound from below that of the walk on P_{n+1}. But from our previous results the edge cover time for a walk on P_n is maximized by a walk started at vertex $n/2 - 1$. Thus we have

$$E_{\frac{n}{2}-1} C^e_{P_n} \leq E_{\frac{n}{2}-1} C^e_{P_{n+1}} < E_{\frac{n}{2}+1} C^e_{P_{n+1}}.$$

If we denote by A_n the expected additional number of steps to complete the edge cover after visiting all vertices on P_n then using (3.28)

$$\begin{aligned} A_n &< \left(\frac{5}{4}\right) n^2 - \left(\frac{5}{4}\right)(n-1)^2 + \frac{9}{4} + A_{n+1} \\ A_n &< \left(\frac{5}{2}\right) n + \frac{7}{2} + A_{n+1}. \end{aligned}$$

Since we have just proved that $A_{n+1} = O(n^2/\log n)$ we now have $A_n = O(n^2/\log n)$ as desired. ∎

CHAPTER IV

GENERAL RESULTS FOR EDGE COVER TIME ON TREES

4.1 Minimum Edge Cover Time for a Tree

In Chapter 1 we mentioned a number of results that have been proven for the vertex cover time for trees. Having extensively studied one tree—the path, we now focus on other trees and show parallel results for edge cover time. The main result of this chapter is the theorem below.

Theorem 4.1 *For any $n \geq 3$, the star on n vertices, S_n, with starting vertex at the center, has the minimum edge cover time among all trees on n vertices.*

In order to prove the theorem we will first show that the minimum edge cover time for any graph is attained by a walk started from some non-leaf vertex. We will then prove that the edge cover time of a tree must be at least 2 more steps than the *minimum* vertex cover time for that tree and finally argue that the star is the only graph which achieves this minimum.

Having proved this theorem we will also determine the probability that a random walk on S_n completes a vertex cover in exactly k steps which leads to two interesting identities involving the Stirling Numbers of the Second Kind. We will note that the edge cover time for a walk on any tree is bounded above by the sum of the vertex cover time and the maximum hitting time in that tree. This will lead to a bound on the edge cover time of a balanced k-ary tree of $O(n\log^2 n + n\log n)$. Finally considering all we have established regarding the edge cover time of random walks on trees we conjecture that it is maximized by a walk started at the center/centers of a path.

While studying edge cover walks on the path we determined that the edge cover time for a random walk on P_n, started at an endpoint, exceeded that for a walk started at its point of attachment by exactly $(n-2)/(n-1)$ steps. It turns out this is true for the star as well. That is, the edge cover time for a walk started from a leaf on S_n exceeds that of a walk started at the center by $(n-2)/(n-1)$ steps. This is the reverse of the situation for the vertex cover in which the vertex cover time is less for walks started at a leaf than for walks started at their points of attachment. These results led us to ask the question: For any graph, G, can we characterize the edge cover time for a walk started from a leaf of G in terms of the edge cover time for a walk started at its point of attachment. The answer is yes.

Lemma 4.2 *Given a graph, G, with at least 3 vertices, a vertex, x, of G with degree 1, and vertex, y, the neighbor of x in G, then*

$$E_x C^e =$$

$$E_y C^e + 1 - \Pr[\textit{walk started at } x \textit{ completes an edge cover on edge } (\text{y,x})].$$

Proof

Consider the set of all walks started at y and observed until they first achieve an edge cover of G. This set may be partitioned into two classes based on the last edge the walk traverses.

A_y = walks started at y which complete an edge cover of G on edge (x, y).

B_y = walks started at y which complete an edge cover of G on an edge other than (x, y).

Likewise we may partition the set of all walks started at x and observed until they first achieve an edge cover of G into two classes.

$$A_x \quad = \quad \text{walks started at x which complete an edge cover of G on edge } (y, x).$$

$$B_x \quad = \quad \text{walks started at x which complete an edge cover of G on an edge other than } (y, x).$$

Given an element of set A_y one may create a unique element of set A_x by adding a step from x to y to the beginning of the walk and deleting the step from x to y from the end of it. Similarly one may take an element of A_x and create a unique element of A_y by the reverse process. Thus there is a bijection between sets A_y and A_x so they must have the same cardinality, and moreover, the operation described does not change the length of the walks so the expected length of the walks in each set must be the same.

Now consider the walks in B_y and B_x. Each walk in B_y may be converted to a walk in B_x by the addition at the beginning of the walk of a step from x to y. One may reverse this process to convert any element of B_x into an element of B_y. Thus these sets also have the same cardinality and the expected length of the walks in B_y must be 1 step less than the expected length of the walks in B_x.

If we indicate the expected length of a walk in X by L_X, then putting these observations together yield

$$\begin{aligned} E_y C^e &= P(A_y) L_{A_y} + P(B_y) L_{B_y} \\ E_x C^e &= P(A_x) L_{A_x} + P(B_x) L_{B_x} \end{aligned}$$

But $P(A_x) = P(A_y) > 0, P(B_x) = P(B_y) > 0, L_{A_x} = L_{A_y}$ and $L_{B_x} = L_{B_y} + 1$ so

$$E_x C^e = E_y C^e + P(B_y) = E_y C^e + 1 - P(A_x). \qquad \blacksquare$$

Thus from Lemma 4.2 we now know that the minimum expected time to cover all the edges of a graph occurs from some non-leaf vertex. In particular, since S_n has only one non-leaf vertex, it is easy to compute the minimum edge cover time for it. An edge cover of the star for a walk started at the center consists of one step to some leaf followed by a vertex cover, and then one additional step to traverse the edge from the vertex at which the vertex cover ended back to the center. Thus the minimum edge cover time on S_n is exactly $E_z C^v + 2$ where z is any leaf of S_n.

Proof of Theorem 4.1

Let G be any tree on $n > 4$ vertices, and suppose that y is a non-leaf vertex of G which minimizes the edge cover time on G. Brightwell and Winkler[7] point out that the vertex cover time from y exceeds the vertex cover time starting from some leaf in G by at least one step, with equality only when G is S_n. This is true because a walk which completes a vertex cover of G creates a spanning tree of G in the process. If such a walk is started at a non-leaf vertex it must walk to some leaf and then trace out the tree. The star is the only tree in which, starting from a non-leaf vertex, the expected time to reach any leaf is exactly 1 step. Thus if G is any tree other than S_n, $E_yC^v > E_xC^v + 1$ for some leaf x. An edge cover in G started at y will consist of a vertex cover and then at least 1 more step. Thus

$$E_yC^e \geq E_yC^v + 1 \geq E_xC^v + 2,$$

with equality on the right only when G is S_n.

Brightwell and Winkler proved that the minimum vertex cover time for any tree on n vertices occurs for a walk on S_n started from a leaf. Thus, if G is not S_n we have

$$E_yC_G^e \geq E_xC_G^v + 2 > E_zC_{S_n}^v + 2,$$

where z is any leaf of S_n.

This leaves only the cases where $n \leq 4$. The only tree on 3 vertices is the star. For $n = 4$ there are two trees: the path and the star. Using the results of Chapter 3 the minimum edge cover time for P_4 is found to be 12 3/4 steps for a walk started at one of the centers, while the minimum on S_4 is 11 steps for walks started at the center. Thus the theorem is true $\forall n \geq 3$. ∎

Given the result in Lemma 4.2 it is now easy to determine the maximum edge cover time for S_n.

Corollary 4.3 *The maximum edge cover time on S_n occurs for a random walk started at a leaf and has value*

$$2(n-1)h_{n-2} + 2 + \left(\frac{n-2}{n-1}\right). \tag{4.1}$$

From Lemma 4.2 we know that this maximum occurs from a leaf, say it is labeled x, and that the edge cover time from a leaf exceeds that from the center, say it is labeled y, by

$$1 - \Pr[\textit{walk started at } x \textit{ completes an edge cover on edge } (\text{y,x})].$$

Proposition 4.4 *The probability that a random walk on S_n, started at a leaf, x, completes a vertex cover without returning to x is $\left(\frac{1}{n-1}\right)$.*

Proof of Proposition 4.4

Consider the sequence of leaves visited by such a walk. Having started at x the walk visits the center point and is then equally likely to step to any of the $n-1$ leaves. Now consider all random walks started from the center. It is easy to see that the probability any such walk ends at vertex x is $1/(n-1)$. Thus of all walks started at x, $1/(n-1)$ will fail to revisit x prior to completing a vertex cover. ∎

This result proves that the probability that a random walk on S_n started at a leaf, x, completes an edge cover on the edge from the center to x is $1/(n-1)$. Thus we have the result in Corollary 4.3. ■

4.2 Stirling Number Identities From Stars

There is a second argument proving Proposition 4.4 which, while more technical in nature, provides two interesting identities involving the Stirling Numbers of the Second Kind. First we will need to determine the probability that a random walk started at a leaf of S_n completes a vertex cover exactly on step $2k$, k an integer.

Lemma 4.5 *Given a random walk on S_n started at some leaf, x, the probability that this random walk completes a vertex cover on step $m = 2k$ is given by*

$$P(E_x C^v = 2k) = \frac{(n-2)!S(k, n-2)}{(n-1)^k}, \tag{4.2}$$

where $S(a, b)$ represents a Stirling number of the Second Kind.

Proof

Let $m = 2k$. Note that k must be at least $n-2$. For this argument consider the center vertex to be labeled y. The vertices visited by the random walk form a sequence in which we may consider position 0 to be filled by our start vertex and every odd element will be filled by the character y. Suppose the starting point is chosen with uniform probability from the leaves of S_n. Then the walk will fill $k+1$ slots in the sequence from an "alphabet" of $n-1$ characters. If the walk completes a vertex cover on step m it must have visited all but one of the $n-1$ leaves by step $m-2=2(k-1)$. In order to count the number of sequences that fulfill this requirement we may first identify one of the leaves to be missed prior to step m. There are $n-1$ ways to do this. Then the number of ways to fill k slots in the sequence from an alphabet of size $n-2$, assuming all characters must appear at least once, is $(n-2)!S(k,n-2)$. The total number of ways to fill k slots from an alphabet of $n-1$ characters is $(n-1)^k$. The probability that on step m we reach the remaining unvisited vertex is $1/(n-1)$. Combining these results we have that the probability that a random walk completes a vertex cover on step $m=2k$ is

$$\left(\frac{1}{n-1}\right)\frac{(n-1)(n-2)!S(k,n-2)}{(n-1)^k}=\frac{(n-2)!S(k,n-2)}{(n-1)^k}.$$

But, since the starting vertex was chosen with uniform probability this is also the fraction of all walks starting at vertex x which complete a vertex cover on step m. ■

Since the random walk completes a vertex cover with probability 1 we have

$$\sum_{k=n-2}^{\infty} P[E_x C^v = 2k] = \sum_{k=n-2}^{\infty} \frac{(n-2)!S(k,n-2)}{(n-1)^k} = 1.$$

which leads immediately to the first of our two identities.[1]

Corollary 4.6

$$\sum_{k\geq 0} \frac{S(k,n)}{(n+1)^k} = \frac{1}{n!}.$$

This may be confirmed directly as follows:

$$\begin{aligned}
&\sum_{k=n-2}^{\infty} P[E_x C^v = 2k] = \\
&= \sum_{k=n-2}^{\infty} \frac{(n-2)!}{(n-1)^k} S(k,n-2) \\
&= \sum_{k=n-2}^{\infty} \frac{1}{(n-1)^k} \sum_{j=0}^{n-2} \binom{n-2}{j} (-1)^j (n-j-2)^k \\
&= \sum_{j=0}^{n-2} \binom{n-2}{j} (-1)^j \sum_{k=n-2}^{\infty} \left(\frac{n-j-2}{n-1}\right)^k \\
&= \sum_{j=0}^{n-2} \binom{n-2}{j} (-1)^j \left(\frac{n-j-2}{n-1}\right)^{n-2} \left(\frac{n-1}{j+1}\right) \\
&= \sum_{j=0}^{n-2} \binom{n-1}{j+1} (-1)^j \left(\frac{(n-1)-(j+1)}{n-1}\right)^{n-2} \\
&= -\sum_{r=1}^{n-1} \binom{n-1}{r} (-1)^r \left(\frac{(n-1)-r}{n-1}\right)^{n-2}
\end{aligned}$$

[1]Wilf points out that these identities may also be proven by appropriate substitutions for x in the generating function $\sum_{k\geq 0} S(k,n)x^k = x^n/(1-x)(1-2x)\cdots(1-nx)$.

$$
\begin{aligned}
&= -\sum_{r=0}^{n-1} \binom{n-1}{r}(-1)^r \left(\frac{(n-1)-r}{n-1}\right)^{n-2} + \left(\frac{n-1}{n-1}\right)^{n-2} \\
&= -\left(\frac{(n-1)!S(n-2,n-1)}{(n-1)^{n-2}}\right) + 1 = 1.
\end{aligned}
$$

The last equation follows from the fact that $S(n-2, n-1) = 0$.

We may use similar reasoning to calculate what fraction of random walks complete a vertex cover on step m and have not returned to the start vertex. We may choose the vertex which will be visited on step m in $n-2$ ways. Since slot 0 is assumed to be filled by the character x, and none of the remaining slots may be filled by x, or the end vertex, we must fill $k-1$ slots from an "alphabet" of $n-3$ characters using each character at least once. This can be done in $(n-3)!S(k-1, n-3)$ ways. Thus the total number of strings matching our requirements is

$$(n-2)(n-3)!S(k-1, n-3).$$

There are $(n-1)^{k-1}$ strings starting at x and ending at step $m-2$. The probability we step to the remaining unvisited vertex on step m is still $1/(n-1)$. So the probability that the random walk completes a vertex cover without returning to the starting point is

$$\left(\frac{1}{n-1}\right)\frac{(n-2)(n-3)!S(k-1,n-3)}{(n-1)^{k-1}} = \frac{(n-2)!S(k-1,n-3)}{(n-1)^k}.$$

From our earlier argument we know that the probability a random walk on the star completes a vertex cover without returning to the starting point is $1/(n-1)$. Thus we obtain

$$\sum_{k=n-2}^{\infty} \frac{(n-2)!S(k-1,n-3)}{(n-1)^k} = \left(\frac{1}{n-1}\right).$$

which leads to the second identity.

Corollary 4.7

$$\sum_{k\geq 0} \frac{S(k,n-2)}{n^{k+1}} = \frac{1}{n!}.$$

This identity may be established directly by remembering that

$$S(x,y) = yS(x-1,y) + S(x-1,y-1).$$

When applied to (4.2) we have

$$\begin{aligned}
1 &= \sum_{k=n-2}^{\infty} \frac{(n-2)!S(k,n-2)}{(n-1)^k} \\
&= \sum_{k=n-2}^{\infty} \frac{(n-2)!}{(n-1)^k} \left[(n-2)S(k-1,n-2) + S(k-1,n-3)\right] \\
&= \sum_{k=n-2}^{\infty} \frac{(n-2)!}{(n-1)^k} (n-2)S(k-1,n-2) \\
&\quad + \sum_{k=n-2}^{\infty} \frac{(n-2)!}{(n-1)^k} S(k-1,n-3).
\end{aligned}$$

Thus

$$\begin{aligned}
&\sum_{k=n-2}^{\infty} \frac{(n-2)!}{(n-1)^k} S(k-1,n-3) = \\
&\quad = 1 - \sum_{k=n-2}^{\infty} \frac{(n-2)!}{(n-1)^k} (n-2)S(k-1,n-2)
\end{aligned}$$

$$
\begin{aligned}
&= 1 - \left(\frac{n-2}{n-1}\right) \sum_{k=n-2}^{\infty} \frac{(n-2)!}{(n-1)^{k-1}}\, S(k-1, n-2) \\
&= 1 - \left(\frac{n-2}{n-1}\right) \sum_{k=n-1}^{\infty} \frac{(n-2)!}{(n-1)^{k-1}}\, S(k-1, n-2) \\
&= 1 - \left(\frac{n-2}{n-1}\right) \sum_{j=n-2}^{\infty} \frac{(n-2)!}{(n-1)^{j}}\, S(j, n-2) \\
&= 1 - \left(\frac{n-2}{n-1}\right) = \frac{1}{n-1}.
\end{aligned}
$$

4.3 A Global Upper Bound for Trees

In Section 2.2 we established an upper bound on edge cover time of $O(dB_\epsilon)$. In Chapter 3 we proved that while this bound would provide $O(2n^2)$ on P_n, in reality the edge cover time is bounded by $(5/4)n^2 + O(n^2/\log n)$. In this section we establish an upper bound specific to trees which improves on $O(dB_\epsilon)$ in many cases.

Inspiration for the following theorem comes from a frequently used property of hitting time known as the *Essential Edge Lemma.* It makes computation of the hitting time between two vertices connected by a unique path relatively straightforward. The form below is due to Brightwell and Winkler[7].

Lemma 4.8 *Let vertices x and y be at distance k in a graph G, with a unique path $x = v_0, v_1, \ldots, v_k = y$ between them. For each $i, 0 < i < k$, let G_i be the* *component of* $G - v_{i-1} - v_{i+1}$ *that contains the point* v_i*; similarly* G_0 *will be the component of* $G - v_1$ *containing* x*. Let* m_i *be the number of edges in* G_i*. Then the expected hitting time* $E_x H_y$ *is equal to*

$$k^2 + 2\sum_{i=0}^{k-1} m_i(k-i).$$

Of course, in a tree all vertices are connected by a unique path. Thus we wondered if this property could be exploited in computation of the edge cover time as well.

Theorem 4.9 *Given a tree,* T*,*

$$C_T^e \leq C_T^v + H_T. \tag{4.3}$$

Proof

Let x be a vertex in T for which $E_x C^e$ is maximized. Consider T to be rooted at x. Consider the forest created when x is deleted from T and call each of the subtrees of T in this forest a branch of T. Then an edge cover walk on T must proceed from the starting point x and first accomplish a vertex cover of T. While accomplishing this vertex cover the walk must visit each of the leaves in each of the branches of T. The vertex cover walk is completed at a leaf in some branch of T. At this point it is the case that the walk has traversed every directed edge in all other branches of T. Thus the only directed edges which may not have been traversed are the "back" edges on the unique path from the leaf at which the vertex cover was completed to the root x. Thus the expected number of steps remaining to complete an edge cover is no more than the hitting time from this leaf to x providing the desired result. ■

This result gives an immediate bound for the edge cover time of balanced k-ary trees on n vertices. Zuckerman[26] proves that the vertex cover time for such trees is $\Theta(n \log^2 n)$. Aldous[2] points out that using the Essential Edge Lemma one may prove that the maximum hitting time in such trees is $O(n \log n)$.

Corollary 4.10 *The edge cover time for a balanced k-ary tree on n vertices is $\Theta(n \log^2 n)$.*

Combining the results of Zuckerman and Aldous with Theorem 4.9 provides the proof of this corollary.

Thus we have another example of a tree in which the expected additional number of steps needed, after completing a vertex cover, in order to assure an edge cover is of lower order than the vertex cover time. Note however, that applying this result to P_n would not have provided the correct result since the vertex cover time and the maximum hitting time on the path are both $\Theta(n^2)$.

When one considers Brightwell and Winkler's conjecture that vertex cover time for trees is maximized on the path, in conjunction with Theorem 4.9, Corollary 4.10 and what we've established for the edge cover time on the path, it seems reasonable to believe that edge cover time is maximized for trees by a walk started from the center of a path. This is certainly one question demanding further study.

CHAPTER V

A TIGHT BOUND ON THE EDGE COVER TIME OF THE CYCLE

5.1 Vertex Cover Time on C_n

A cycle on n vertices, C_n, is a connected, regular, simple graph in which each vertex has degree 2. We briefly mentioned in Section 1.2 that the vertex cover time for C_n is $\binom{n}{2}$. One classic argument establishing this is a combination of a coupon collectors approach with the recursive argument used to establish the hitting times on the path. The outline here parallels Wilf[23].

Consider the point in the random walk on C_n at which the j^{th} new vertex has first been visited. Then the walk has visited j contiguous vertices which form a path and is resting on the endpoint of the path. How long will it take, on average, to visit the $(j+1)^{st}$ vertex? We may consider the walk to be on a path with $j+2$ vertices having covered the vertices from 2 to $j+1$ and resting at vertex $j+1$. We then want to know: how long, on average, before the walk visits vertex 1 or vertex $j+2$, whichever occurs first? If we let c_r represent the expected number of steps, starting at vertex r to first visit either vertex 1 or $j+2$ on such a path, then $c_1 = c_{j+2} = 0$ and for $1 < k < j+2$,

$$c_k = \left(\frac{1}{2}\right)(c_{k-1}+1) + \left(\frac{1}{2}\right)(c_{k+1}+1).$$

The solution to this recurrence relation is $c_r = r - r^2 + (r-1)(j+2)$. Which provides $c_{j+1} = j$.

Now the vertex cover time is easily computed.

$$C^{v}_{C_n} = \sum_{j=1}^{n-1} j = \binom{n}{2}. \qquad \blacksquare$$

Due to the symmetry of the cycle this, as well as all edge cover times, is independent of the starting vertex. From now on we will assume the cycle is labeled in the natural way and that all walks begin at vertex 1.

5.2 The Undirected Edge Cover Time of C_n

Path results may also be used to compute the edge cover times for cycles. In the case of the undirected edge cover time we do nothing more than extend the connection above. In the directed case we need to use the more delicate results established in Chapter 3. As we have done in the previous cases we will calculate the edge cover time by first finding the vertex cover time and then analyzing the expected number of additional steps needed to complete an edge cover. In Chapter 3 we pointed out that the undirected edge cover time of a tree is the same as the vertex cover time. This is not the case for the cycle.

Theorem 5.1 *The undirected edge cover time for* C_n *is*

$$\binom{n+1}{2}.$$

Proof

Upon completing a vertex cover of C_n, say at vertex j, the only undirected edge not yet traversed is either $\{j, j+1\}$ or $\{j, j-1\}$. The expected number of additional steps needed in order to traverse this last edge may be calculated by considering a walk on a path containing $n+2$ vertices. The walk on the cycle may be considered to have been a walk which has visited vertices $\{2, 3, \ldots, n, n+1\}$ of this path and which is now located at vertex $n+1$. Then if we compute the expected amount of time necessary for such a walk to either visit vertex 1 or vertex $n+2$, whichever occurs first, then this will be the expected number of steps necessary to traverse the remaining edge on the cycle. The argument above shows that it takes n steps to reach either vertex 1 or vertex $n+2$ of the path. Combining this with the vertex cover time for C_n we have

$$C^{ue} = C^v + n = \binom{n}{2} + n = \binom{n+1}{2}. \qquad \blacksquare$$

5.3 The Directed Edge Cover Time for C_n

Lovász and Winkler[19] point out that it is "folklore" that the probability a particular vertex $j \neq 1$ is the last new vertex visited by a random walk started at 1 on C_n is $1/(n-1)$. The key to analyzing the directed edge cover time is to realize that upon completing the vertex cover portion of the walk the sequence of vertices visited may be mapped to a sequence describing a vertex cover of P_n with the starting point determined by the vertex at which the walk completes a vertex cover of C_n. Using this fact we will be able to relate the probability distribution for untraversed edges on the path to that of edges on the cycle. We will then use this to determine lower and upper bounds on the expected number of additional steps needed to complete an edge cover after all the vertices have been visited at least once.

Theorem 5.2 *The directed edge cover time for a simple random walk on C_n is*

$$\binom{n}{2} + \Theta\left(\frac{n^2}{\log n}\right).$$

Proof

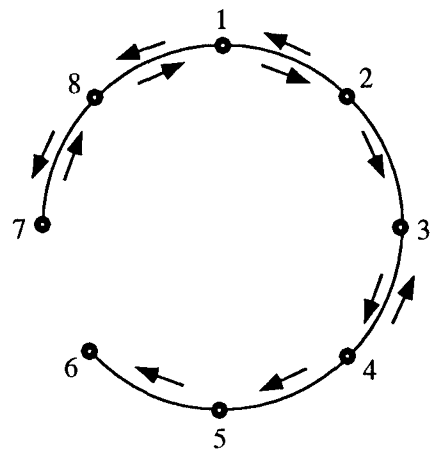

Figure 5.1: Vertex cover walk of C_8

Suppose we label P_n and C_n in the natural way. Also suppose that we take a walk on C_n and it completes a vertex cover at vertex $2 \leq j \leq n$ having stepped to j from vertex $j-1$. The probability that this occurs may be determined by a simple probability tree analysis which yields $(j-1)/n(n-1)$.

If we map vertex $(j+k) mod\, n$ on C_n to vertex k on P_n, so that 1 on C_n maps to $n-j+1$ on P_n and j on C_n maps to n on P_n, then the sequence of vertices visited on C_n will correspond to a vertex cover walk on P_n started at vertex $n-j+1$ and ending at vertex n. A vertex cover walk on C_8 is illustrated in figure 5.1, and the mapping to a vertex cover walk on P_8 is illustrated in figure 5.2. As a matter of fact this mapping describes a bijection between vertex cover walks starting at vertex 1 and ending at j on C_n and vertex cover walks on P_n starting at vertex $n-j+1$ and ending at vertex n.

Figure 5.2: Vertex cover walk from figure 5.1 mapped to vertex cover walk of P_8 ending at 8

Now consider how this walk on C_n must have visited the vertices in the cycle. In order to end a vertex cover at j stepping to it from $j-1$ the walk must have traveled back and forth to either side of the starting vertex until it visited vertex $j+1$. It then traveled back and forth, never proceeding beyond $j+1$ until it finally visited j for the first time from $j-1$. This means that all edges between vertex $j+1$ and vertex 1 have been traversed in both directions as illustrated by the walk in figure 5.1. Thus the only possible edges not traversed during the vertex cover portion of the walk are directed edges $(j, j+1)$, $(j+1, j)$, $(j, j-1)$, $\ldots$, $(2, 1)$. Given the bijection between walks on the cycle and walks on the path described above we can describe the probability that t is the least integer such that edge $(t, t-1)$ is the closest edge to vertex 1 not traversed during the vertex cover portion of the walk on C_n, for $2 \leq t \leq j-1$. Using the notation of Section 3.4 this probability is $P(X^n_{n-j+1,n-j+t} \mid A^{1\to n}_{n-j+1})$.

If edge $(t, t-1)$ is the edge closest to vertex 1 not traversed so far, then we may obtain an upper bound for the expected number of steps remaining to complete an edge cover by the expected amount of time to traverse the edges $(j, j+1)$, $(j+1, j)$, $(j, j-1)$, $(j-1, j-2), \ldots$, $(t+1, t)$, $(t, t-1)$ in this particular order. Once again a recursive argument may be used to show that the expected number of steps needed to traverse edge $(s, s+1)$ starting at s on C_n is $n+1$. It follows that the expected number of steps to traverse these edges in the specified order is $(j-t+3)(n+1)$.

Thus, given that edge $(t, t-1), 2 \leq t \leq j-1$, is the closest edge to vertex 1 not yet traversed, an upper bound on the expected number of remaining steps is

$$(j-t+3)(n+1)P(X^n_{n-j+1,n-j+t} \mid A^{1\to n}_{n-j+1}). \tag{5.1}$$

If the closest edge to vertex 1 not traversed was $(j, j-1)$, then an upper bound on the expected number of steps remaining is $3(n+1)$ and the probability that this occurs is less than 1.

Putting this together with the probability that a walk does end a vertex cover at j by stepping from $j-1$, the additional number of steps needed is bounded above by

$$\sum_{j=2}^{n} \left(\frac{j-1}{n(n-1)}\right) \Bigg[3(n+1) + \sum_{t=2}^{j-1} (j-t+3)(n+1) P(X^{n}_{n-j+1,n-j+t} \mid A^{1\to n}_{n-j+1})\Bigg]. \qquad (5.2)$$

Remembering the symmetry of C_n, this also describes a bound on the additional number of steps needed for walks ending a vertex cover at j by stepping from $j+1$. Thus an upper bound on the expected additional number of steps needed to complete an edge cover, having visited all vertices at least once, is twice the quantity in (5.2). Denote this value by A_n.

Using equation(3.23) yields

$$\begin{aligned} A_n &= 2\sum_{j=2}^{n} \left(\frac{j-1}{n(n-1)}\right) \Bigg[3(n+1) + \sum_{t=2}^{j-1} (j-t+3)(n+1) \left(\frac{n-1}{j-1}\right) \times \\ &\qquad \left(\frac{1}{j-t+1}\right) \left(\frac{t-1}{n-j+t-1}\right) P(X^{n-j+t}_{n-j+1,n-j+t} \mid A^{1\to n-j+t}_{n-j+1})\Bigg] \\ &= 6\left(\frac{n+1}{n(n-1)}\right) \sum_{j=2}^{n} (j-1) \\ &\quad + 2\left(\frac{n+1}{n}\right) \sum_{j=2}^{n}\sum_{t=2}^{j-1} \left(\frac{j-t+3}{j-t+1}\right) \left(\frac{t-1}{n-j+t-1}\right) \times \\ &\qquad\qquad P(X^{n-j+t}_{n-j+1,n-j+t} \mid A^{1\to n-j+t}_{n-j+1}) \\ &\le 3(n+1) + 4\left(\frac{n+1}{n}\right) \sum_{j=2}^{n}\sum_{t=2}^{j-1} \left(\frac{t-1}{n-j+t-1}\right) \times \\ &\qquad\qquad P(X^{n-j+t}_{n-j+1,n-j+t} \mid A^{1\to n-j+t}_{n-j+1}). \qquad (5.3) \end{aligned}$$

Note that since the first term in this sum is $o(n^2/\log n)$ we need only prove that the second term is $O(n^2/\log n)$. Also note that in the sum in this inequality the term with $j = 2$ is zero, so the sum for which we desire a bound is

$$S = 4\left(\frac{n+1}{n}\right)\sum_{j=3}^{n}\sum_{t=2}^{j-1}\left(\frac{t-1}{n-j+t-1}\right)P(X_{n-j+1,n-j+t}^{n-j+t} \mid A_{n-j+1}^{1\to n-j+t}). \quad (5.4)$$

Using equation(3.26) (5.4) becomes

$$\begin{aligned}
S &= 4\left(\frac{n+1}{n}\right)\sum_{j=3}^{n}\sum_{t=2}^{j-1}\sum_{r=1}^{t-1}(f_r - f_{r-1})\left(\frac{t-r}{n-j+t-r}\right). \\
&= 4\left(\frac{n+1}{n}\right)\sum_{j=3}^{n}\sum_{r=1}^{j-2}(f_r - f_{r-1})\sum_{t=r+1}^{j-1}\left(\frac{t-r}{n-j+t-r}\right) \\
&= 4\left(\frac{n+1}{n}\right)\sum_{j=3}^{n}\left[\sum_{r=1}^{j-2} f_r \sum_{k=1}^{j-r-1}\left(\frac{k}{n-j+k}\right)\right. \\
&\qquad \left. - \sum_{r=1}^{j-3} f_r \sum_{k=1}^{j-r-2}\left(\frac{k}{n-j+k}\right)\right] \\
&= 4\left(\frac{n+1}{n}\right)\sum_{j=3}^{n}\sum_{r=1}^{j-2} f_r\left(\frac{j-r-1}{n-r-1}\right) \\
&= 4\left(\frac{n+1}{n}\right)\sum_{r=1}^{n-2}\left(\frac{f_r}{n-r-1}\right)\sum_{k=1}^{n-r-1} k \\
&= 2\left(\frac{n+1}{n}\right)\sum_{r=1}^{n-2}(n-r)f_r \\
&< 2\left(\frac{n+1}{n}\right)\sum_{r=1}^{n-1}(n-r)f_r.
\end{aligned}$$

But by definition (3.8) this gives

$$S = 2\left(\frac{n+1}{n}\right)g_{n-1}. \quad (5.5)$$

Thus

$$A_n \leq 3(n+1) + 2\left(\frac{n+1}{n}\right)g_{n-1},$$

and in Lemma 3.4 we have shown that $g_{n-1} = O(n^2/\log n)$. Therefore the expected number of steps needed to complete an edge cover after first visiting all the vertices on C_n is $O(n^2/\log n)$.

This upper bound was shown based on traversing the remaining directed edges in a specific order. We may obtain a lower bound on the additional number of steps needed by determining the edge closest to vertex 1 which has not been traversed, and then computing the hitting time from our present position back to traverse that edge. Simple recursive arguments show that the hitting time from vertex s to traverse edge $(t, t-1), 2 \le t \le s-1$ is $(s-t+1)(n-s+t+1)$ steps. Thus a lower bound on the expected additional number of steps to complete an edge cover is

$$\begin{aligned} A_n &\geq 2\sum_{j=3}^{n}\sum_{t=2}^{j-1}(j-t+1)(n-j+t+1)\left(\frac{j-1}{n(n-1)}\right)\binom{n-1}{j-1}\times \\ &\qquad \left(\frac{1}{j-t+1}\right)P(X_{n-j+1,n-j+t}^{n-j+t} \mid A_{n-j+1}^{1\to n-j+t}). \end{aligned} \tag{5.6}$$

Once again using equation(3.26), and simplifying, (5.6) becomes

$$\begin{aligned} A_n &\geq 2\sum_{j=3}^{n}\sum_{t=2}^{j-1}\left(\frac{n-j+t+1}{n}\right)\sum_{r=1}^{t-1}(f_r - f_{r-1})\left(\frac{t-r}{n-j+t-r}\right) \\ &\geq \left(\frac{2}{n}\right)\sum_{j=3}^{n}\sum_{t=2}^{j-1}\sum_{r=1}^{t-1}(f_r - f_{r-1})(t-r) \\ &= \left(\frac{2}{n}\right)\sum_{j=3}^{n}\sum_{t=2}^{j-1}\sum_{r=1}^{t-1} f_r. \end{aligned}$$

Using Corollary A.5 we then have

$$
\begin{aligned}
A_n &\geq \left(\frac{2}{n}\right)\sum_{j=3}^{n}\sum_{t=2}^{j-1}\left(\frac{1}{2}\right)\frac{t-1}{h_{t-1}} \\
&\geq \left(\frac{1}{n}\right)\sum_{j=3}^{n}\frac{1}{h_{j-2}}\sum_{t=1}^{j-2} t \\
&\geq \left(\frac{1}{2n}\right)\sum_{j=1}^{n-2}\left(\frac{1}{h_j}\right)(j)(j+1) \\
&\geq \left(\frac{1}{2nh_{n-2}}\right)\sum_{j=1}^{n-2} j^2 \\
&= \Omega\left(\frac{n^2}{\log n}\right).
\end{aligned}
$$

Thus we have the desired result that $A_n = \Theta(n^2/\log n)$ proving that $C^e_{C_n} = \binom{n}{2} + \Theta(n^2/\log n)$. ∎

CHAPTER VI

QUESTIONS FOR FURTHER STUDY

We established that S_n minimizes edge cover time for trees. Which class of graphs minimizes edge cover time for all graphs? Which maximizes it?

Our results for edge cover time on trees seem to suggest that, for fixed n, the more leaves a tree contains the smaller the number of steps needed to complete a vertex cover walk to an edge cover walk. If one could prove Brightwell and Winkler's conjecture that vertex cover time for trees is maximized by the path, then this observation would seem to indicate the same could be proven for edge cover time.

Brightwell and Winkler[6] established that the maximum hitting time for any graph on n vertices is obtained by a walk on the "lollipop" and has order $(4/27)n^3 + o(n^3)$. Feige went on to prove that the vertex cover time for any graph on n vertices is $(4/27)n^3 + O(n^{2.5})$, which makes the lollipop graph a candidate for the graph which maximizes vertex cover time. Since

Zuckerman's bound proves that edge cover time is $O(n^3)$ it seems plausible that the "lollipop" may maximize edge cover time as well. Is there a way to prove this? Simulations on this graph, for small n, seem to indicate that the expected additional number of steps needed to complete a vertex cover to an edge cover is $O(n^2)$. Can this be proved, perhaps using methods similar to those used to establish our global upper bound?

Bounds for vertex cover time tighter than those developed by Matthews and Aleliunas have been established based on knowing the maximum degree, average degree, or that the graph is regular. Can improved bounds be derived for edge cover time based on these parameters?

The blanket time provides a measure of when the walk has "distributed" itself evenly amongst the vertices of the graph. If we define a directed and undirected "edge blanket time" in a similar manner, can we establish bounds on it?

Winkler and Zuckerman[25] use ad hoc methods to prove that the blanket time for the path and cycle is $O(C^v)$. Their results only provide a bound on the constant of e^{48}. Intuitively it seems to us that the constant in both cases should be small. This is because a walk on the path started at an endpoint spends much of its time in the "first half" of the path and very little in the "second half". Thus if a walk is allowed to complete say 4 vertex covers then it seems likely that it has distributed itself evenly over both ends of the path and the only possible area of difficulty would be the number of visits to the vertices in the center. However, computer simulations, using code which has proved very accurate in our study of vertex and edge cover time, shows the ratio $B_{1/2}/C^v$ slowly growing from approximately 2.4 for a path of 10 vertices to 4.9 for a path on 100 vertices. In the case of a cycle the ratio grows from 2.7 for a cycle on 5 vertices to 6.2 for a cycle on 200 vertices. The simulations do not give any evidence that this growth is ending. If the results of Winkler and Zuckerman are correct, then what is the correct constant? Is it possible that there is no constant and in fact the ratio is unbounded?

We have established tight bounds on the edge cover time for a number of trees and the cycle. Aldous' and Zuckerman's results combine to provide a tight bound for the clique. Can our techniques be used to provide tight bounds on other classes of graphs?

Many of the results for vertex cover time have been generalized to cases in which each edge is given a weight and the probability of traversing any particular edge out of a vertex is not uniform. Is there a way to do this in the edge cover time case applying our results?

APPENDIX A

PROPERTIES OF $\{f_n\}$

In Sections 3.1-3.3 we considered random walks on P_n started from an endpoint. We assumed the path was labeled in the natural way with the start point being vertex 1 and the other endpoint being vertex n. For each edge $(k, k-1)$, $2 \leq k \leq n$, we determined that the probability the edge was not traversed prior to the walk completing a vertex cover was $1/(n-k+1)$. Given that edge $(k, k-1)$ was not traversed prior to the completion of the vertex cover, we then calculated the probability that it was the first such edge and labeled that probability f_{k-1}. In this section we collect a number of facts about the sequence $\{f_n\}$ which have proved useful in our analysis of the edge cover time of random walks on graphs.

After defining the f_n in Chapter 3 we noted that they fulfilled identity(3.5). This gave us a recursive method for computation of $\{f_n\}$, and made it clear that this is a sequence of rational numbers. Using *Mathematica* it was an easy task to compute the values of the sequence up to f_{750}. Table(A.1) contains

f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}
1	$\frac{1}{2}$	$\frac{5}{12}$	$\frac{3}{8}$	$\frac{251}{720}$	$\frac{95}{288}$	$\frac{19087}{60480}$	$\frac{5257}{17280}$	$\frac{1070017}{3628800}$	$\frac{25713}{89600}$

Table A.1: Exact values of $f_1 - f_{10}$

the values for $f_n, 1 \le n \le 10$.

A second identity, in this case involving a weighted sum of the harmonic numbers, may be derived directly from identity(3.5). Specifically,

$$\begin{aligned} \sum_{j=1}^{n} \sum_{k=1}^{j} \left(\frac{1}{j-k+1} \right) f_k &= \sum_{j=1}^{n} 1, \\ \sum_{k=1}^{n} f_k \sum_{j=k}^{n} \left(\frac{1}{j-k+1} \right) &= n, \\ \sum_{k=1}^{n} f_k h_{n-k} &= n. \end{aligned} \tag{A.1}$$

The connection between the elements of $\{f_n\}$ and the harmonic numbers is pervasive.

The values in Table A.1 clearly suggest that the f_n form a decreasing sequence. This turns out to be the case.

Lemma A.1 *The sequence, $\{f_n\}$, is a strictly decreasing sequence.*

Proof

Using identity(3.5) we have for $n \ge 1$,

$$\sum_{k=1}^{n} \left(\frac{1}{n-k+1} \right) f_k - \sum_{k=1}^{n+1} \left(\frac{1}{n-k+2} \right) f_k = 0.$$

So

$$\sum_{k=1}^{n}\left(\frac{1}{n-k+1}\right)(f_k - f_{k+1}) = \frac{1}{n+1}. \tag{A.2}$$

Writing $s_n = f_n - f_{n+1}$ yields

$$\sum_{k=1}^{n}\left(\frac{1}{n-k+1}\right)s_k = \frac{1}{n+1}. \tag{A.3}$$

Then for $n \geq 2$

$$\begin{aligned}\frac{1}{n+1} - s_n &= \sum_{k=1}^{n-1}\left(\frac{n-k}{n-k+1}\right)\left(\frac{1}{n-k}\right)s_k \\ &< \left(\frac{n-1}{n}\right)\sum_{k=1}^{n-1}\left(\frac{1}{n-k}\right)s_k.\end{aligned}$$

Replacing the sum by (A.3) yields

$$\frac{1}{n+1} - s_n < \frac{n-1}{n^2}.$$

Thus $\forall n \geq 2$,

$$s_n > \frac{1}{n^2(n+1)} > 0. \tag{A.4}$$

Combining this with the fact that $s_1 = 1/2$ we have $\forall n \geq 1$,

$$f_n - f_{n+1} > 0. \qquad \blacksquare$$

Using the fact that the sequence is strictly decreasing we can determine an upper bound on f_n.

Lemma A.2 *For all $n > 2$,*

$$f_n < \left(1 - \frac{1}{2n}\right)\frac{1}{h_n}. \tag{A.5}$$

Proof

Using inequality(A.4) and iterating we have

$$\begin{aligned}
(1 - f_n) &> (1 - f_{n-1}) + \frac{1}{n(n-1)^2} \\
(1 - f_n) &> (1 - f_{n-2}) + \frac{1}{n(n-1)^2} + \frac{1}{(n-1)(n-2)^2} \\
\vdots \quad & \quad \vdots \\
(1 - f_n) &> (1 - f_1) + \sum_{k=0}^{n-2} \frac{1}{(n-k)(n-k-1)^2}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\sum_{k=2}^{n}\left(\frac{1}{k}\right)(1 - f_n) &> \sum_{k=2}^{n}\left(\frac{1}{k}\right)\left((1 - f_{n-k+1}) + \sum_{j=0}^{k-2}\frac{1}{(n-j)(n-j-1)^2}\right) \\
&= (h_n - 1) - (1 - f_n) \\
&\quad + \sum_{j=0}^{n-2}(h_n - h_{j+1})\left(\frac{1}{(n-j)(n-j-1)^2}\right).
\end{aligned}$$

This then implies that

$$-h_n f_n > -1 + \sum_{j=0}^{n-2}(h_n - h_{j+1})\left(\frac{1}{(n-j)(n-j-1)^2}\right).$$

Multiplying this inequality by $-1/h_n$ and discarding all but the last term in the sum on the righthand side we determine that

$$f_n < \left(1 - \frac{1}{2n}\right)\frac{1}{h_n}. \qquad \blacksquare$$

It is worth noting that one can improve on this bound slightly by using equation(3.13). If one integrates this quantity using bounds similar to those in the argument at (3.16) one obtains the bound

$$f_{n+1} < \left(1 - \frac{1}{\alpha n} - \frac{1}{n^2}\right)\frac{1}{h_n} \qquad \forall \alpha > 1.$$

It appears that establishing a lower bound for f_n is not quite so easy.

Lemma A.3

$$f_n = (1 - o(1))\left(\frac{1}{h_n}\right). \tag{A.6}$$

To facilitate obtaining such a bound we introduce a new quantity. Define q_n by

$$q_n = \sum_{k=1}^{n} \left(\frac{1}{n-k+1}\right)\frac{1}{h_k}. \tag{A.7}$$

Claim

$$(q_n - 1)h_n = o(1).$$

Proof

$$\begin{aligned} q_n h_n - h_n &= \left[\sum_{k=1}^{n} \left(\frac{1}{n-k+1}\right)\frac{h_n}{h_k}\right] - h_n \\ &= \left[\sum_{k=1}^{n} \left(\frac{1}{n-k+1}\right)\left(\frac{h_n - h_k}{h_k} + 1\right)\right] - h_n \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{n-1} \left(\frac{1}{n-k+1}\right)\left(\frac{h_n - h_k}{h_k}\right) \\
&= \left(\frac{1}{n}\right)(h_n - 1) + \sum_{k=2}^{n-1} \left(\frac{1}{n-k+1}\right)\left(\frac{h_n - h_k}{h_k}\right). \qquad \text{(A.8)}
\end{aligned}
$$

Since $h_n/n = o(1)$ it suffices to show that the sum is bounded above by a function that approaches 0 as n grows. We will do this by first splitting this sum into three parts. For simplicity of notation let $N = \left\lfloor \frac{n}{\ln^2 n} \right\rfloor$. Since these quantities contain the harmonic numbers we will make frequent use of the inequalities

$$
\ln n + \gamma < h_n < \ln n + \gamma + \frac{1}{2n},
$$

where $\gamma = \lim_{n\to\infty}(h_n - \ln n)$ is *Euler's constant.*

(A.) $\sum_{k=2}^{N} \left(\frac{1}{n-k+1}\right)\left(\frac{h_n-h_k}{h_k}\right)$.

There are less than $\frac{n}{\ln^2 n}$ terms in this sum. For large n, and in the range $2 \leq k \leq N$,

$$
\left(\frac{1}{n-k+1}\right)\left(\frac{1}{h_k}\right) < \frac{1}{n},
$$

and

$$
(h_n - h_k) \leq (h_n - \frac{3}{2}).
$$

Thus

$$
\begin{aligned}
\sum_{k=2}^{N} \left(\frac{1}{n-k+1}\right)\left(\frac{h_n - h_k}{h_k}\right) &\leq \left(\frac{n}{\ln^2 n}\right)\left(\frac{1}{n}\right)(h_n - \frac{3}{2}) \\
&< \frac{1}{\ln n} = o(1).
\end{aligned}
$$

(B.) $\sum_{k=N+1}^{n-N-1} \left(\frac{1}{n-k+1}\right)\left(\frac{h_n-h_k}{h_k}\right)$.

In this sum

$$\begin{aligned}
\left(\frac{h_n - h_k}{h_k}\right) &\leq \left(\frac{h_n - h_{N+1}}{h_{N+1}}\right) \\
&\leq \left(\frac{\ln n + \frac{1}{2n} - \ln(N+1)}{\ln(N+1)}\right) \\
&\leq \left(\frac{\frac{1}{2n} + 2\ln\ln n}{\ln n - 2\ln\ln n}\right). \qquad \text{(A.9)}
\end{aligned}$$

Also

$$\begin{aligned}
\sum_{k=N+1}^{n-N-1} \left(\frac{1}{n-k+1}\right) &\leq h_{n-N} - h_{N+2} \\
&\leq \ln(n-N) + \\
&\quad + \frac{1}{2n\left(1-\frac{1}{\ln^2 n}\right)} - \ln(N+1) \\
&\leq \ln\left(n - \frac{n}{\ln^2 n} + 1\right) \\
&\quad + \frac{1}{2n\left(1-\frac{1}{\ln^2 n}\right)} - \ln\left(\frac{n}{\ln^2 n}\right) \\
&= \ln\left(1 + \frac{1}{n} - \frac{1}{\ln^2 n}\right) \\
&\quad + \frac{1}{2n\left(1-\frac{1}{\ln^2 n}\right)} + 2\ln\ln n \\
&< \frac{1}{2n\left(1-\frac{1}{\ln^2 n}\right)} + 2\ln\ln n. \qquad \text{(A.10)}
\end{aligned}$$

Combining inequalities(A.9) and (A.10) we have

$$\begin{aligned}
\sum_{k=N+1}^{n-N-1} \left(\frac{1}{n-k+1}\right)\left(\frac{h_n - h_k}{h_k}\right) & \\
&\leq \left(\frac{\frac{1}{2n} + 2\ln\ln n}{\ln n - 2\ln\ln n}\right)\sum_{k=N+1}^{n-N-1}\left(\frac{1}{n-k+1}\right) \\
&\leq \left(\frac{\frac{1}{2n} + 2\ln\ln n}{\ln n - 2\ln\ln n}\right)\left(\frac{1}{2n\left(1-\frac{1}{\ln^2 n}\right)} + 2\ln\ln n\right) \\
&= o(1).
\end{aligned}$$

(C.) $\sum_{k=n-N}^{n-1} \left(\frac{1}{n-k+1}\right)\left(\frac{h_n - h_k}{h_k}\right)$.

Here we have

$$\begin{aligned}
\left(\frac{h_n - h_k}{h_k}\right) &\leq \left(\frac{h_n - h_{n-N}}{h_{n-N}}\right) \\
&\leq \left(\frac{\ln n + \frac{1}{2n} - \left(\ln n + \ln\left(1-\frac{1}{\ln^2 n}\right)\right)}{\ln\left(n - \frac{n}{\ln^2 n}\right)}\right) \\
&= \left(\frac{\frac{1}{2n} + \ln\left(\frac{\ln^2 n}{\ln^2 n - 1}\right)}{\ln n + \ln\left(1-\frac{1}{\ln^2 n}\right)}\right).
\end{aligned}$$

However $\ln(x) \leq x - 1$. So

$$\left(\frac{h_n - h_k}{h_k}\right) \leq \frac{\frac{1}{2n} + \frac{1}{\ln^2 n - 1}}{\ln n + \ln\left(1-\frac{1}{\ln^2 n}\right)}.$$

Thus

$$\sum_{k=n-N}^{n-1} \left(\frac{1}{n-k+1}\right)\left(\frac{h_n - h_k}{h_k}\right)$$

$$
\begin{aligned}
&\le \left(\frac{\frac{1}{2n}+\frac{1}{\ln^2 n-1}}{\ln n+\ln\left(1-\frac{1}{\ln^2 n}\right)}\right)\sum_{k=n-N}^{n-1}\left(\frac{1}{n-k+1}\right)\\
&= \left(\frac{\frac{1}{2n}+\frac{1}{\ln^2 n-1}}{\ln n+\ln\left(1-\frac{1}{\ln^2 n}\right)}\right)(h_{N+1}-1)\\
&\le \left(\frac{\frac{1}{2n}+\frac{1}{\ln^2 n-1}}{\ln n+\ln\left(1-\frac{1}{\ln^2 n}\right)}\right)\ln\left(\frac{n}{\ln^2 n}\right)\\
&\le \left(\frac{\frac{1}{2n}+\frac{1}{\ln^2 n-1}}{\ln n+\ln\left(1-\frac{1}{\ln^2 n}\right)}\right)\ln n\\
&= o(1).
\end{aligned}
$$

Combining A, B, and C with equation(A.8) we have the desired result. ■

Proof of Lemma A.3

We know $f_1 = 1/h_1$, $f_2 = (1-1/4)(1/h_2)$ and in Lemma A.2 we established that $\forall n > 2, f_n < (1-1/2n)(1/h_n)$. From identity(3.5) we know that

$$
f_n = 1-\left[\left(\frac{1}{2}\right)f_{n-1}+\left(\frac{1}{3}\right)f_{n-2}+\cdots+\left(\frac{1}{n-1}\right)f_2+\left(\frac{1}{n}\right)f_1\right].
$$

Thus $\forall n \ge 3$

$$
\begin{aligned}
f_n \;&\ge\; 1-\left[\left(\frac{1}{2}\right)\left(\frac{1}{h_{n-1}}-\frac{1}{2(n-1)h_{n-1}}\right)+\left(\frac{1}{3}\right)\left(\frac{1}{h_{n-2}}-\frac{1}{2(n-2)h_{n-2}}\right)+\cdots\right.\\
&\qquad \left.\cdots+\left(\frac{1}{n-1}\right)\left(\frac{1}{h_2}-\frac{1}{2(2)h_2}\right)+\left(\frac{1}{n}\right)\left(\frac{1}{h_1}\right)\right],\\
&\ge\; 1-\left[\left(\frac{1}{2}\right)\frac{1}{h_{n-1}}+\left(\frac{1}{3}\right)\frac{1}{h_{n-2}}+\cdots+\left(\frac{1}{n-1}\right)\frac{1}{h_2}+\left(\frac{1}{n}\right)\frac{1}{h_1}\right]\\
&\qquad +\left(\frac{1}{2}\right)\left[\left(\frac{1}{2}\right)\left(\frac{1}{n-1}\right)\frac{1}{h_{n-1}}+\left(\frac{1}{3}\right)\left(\frac{1}{n-2}\right)\frac{1}{h_{n-2}}+\cdots\right.
\end{aligned}
$$

$$\cdots + \left(\frac{1}{n-1}\right)\left(\frac{1}{2}\right)\frac{1}{h_2}\Bigg].$$

Then subtracting $1/h_n$ from both sides of this inequality yields

$$\begin{aligned} f_n - \frac{1}{h_n} \geq\ & 1 - q_n \\ & + \left(\frac{1}{2}\right)\left[\left(\frac{1}{2}\right)\left(\frac{1}{n-1}\right)\frac{1}{h_{n-1}} + \left(\frac{1}{3}\right)\left(\frac{1}{n-2}\right)\frac{1}{h_{n-2}} + \cdots \right. \\ & \left. \cdots + \left(\frac{1}{n-1}\right)\left(\frac{1}{2}\right)\frac{1}{h_2}\right]. \end{aligned}$$

which leads to the conclusion that

$$f_n > 1 - q_n + \frac{1}{h_n}.$$

Using the conclusion of the claim proved above we have the desired result

$$f_n = (1 - o(1))\,\frac{1}{h_n}. \qquad \blacksquare$$

In Section 3.3 we needed to bound the quantity $g_n = \sum_{j=1}^{n-1} (n-j)f_j$. Using results from Jordan[17] we proved an upper bound of $O(n^2/\log n)$. We also claimed that it is possible to lower bound the quantity by $\Omega(n^2/\log n)$. Below is the proof of that claim.

Lemma A.4

$$g_n = \sum_{j=1}^{n-1} (n-j)f_j = \Omega\left(\frac{n^2}{\log n}\right) \tag{A.11}$$

Proof

The proof will be by induction on the "forward differences" of g_n. $\triangle g_{n-1} = g_n - g_{n-1} = \sum_{k=0}^{n} f_k$ denotes the first forward difference of g_n and $\triangle^{(2)} g_n = \triangle(\triangle g_{n-2}) = f_n$ the second forward difference. As a result of Lemma A.3 we know that $f_n > (1/2)(1/h_n), \forall n \geq 1$. Claim $\triangle g_{n-1} \geq (1/2)(n/h_n)$.

$$\begin{aligned}
\triangle g_1 &= f_1 + f_2 \\
&= \frac{3}{2} \\
&> \left(\frac{1}{2}\right)\frac{4}{3}.
\end{aligned}$$

Now suppose the claim is true $\forall j < n$.

$$\triangle^{(2)} g_{n-2} = \triangle g_{n-1} - \triangle g_{n-2} = f_n$$

implies that

$$\triangle g_{n-1} = \triangle g_{n-2} + f_n.$$

But by the inductive hypothesis

$$\begin{aligned}
\triangle g_{n-2} + f_n &\geq \left(\frac{1}{2}\right)\frac{n-1}{h_{n-1}} + \left(\frac{1}{2}\right)\frac{1}{h_n} \\
&= \left(\frac{1}{2}\right)\frac{(n-1)\frac{1}{n} + (n-1)h_{n-1} + h_{n-1}}{h_n h_{n-1}} \\
&> \left(\frac{1}{2}\right)\frac{n}{h_n}.
\end{aligned}$$

Now that we know $\triangle g_{n-1} \geq \left(\frac{1}{2}\right)\frac{n}{h_n}$ the rest of the induction is straightforward.

$$g_1 = 1 > \frac{1}{4}.$$

Suppose $g_j > \left(\frac{1}{4}\right)\frac{n^2}{h_n}, \forall j < n$. Then

$$g_n - g_{n-1} = \triangle g_{n-1} \geq \left(\frac{1}{2}\right)\frac{n}{h_n}$$

implies

$$\begin{aligned}
g_n &\geq g_{n-1} + \left(\frac{1}{2}\right)\frac{n}{h_n} \\
&\geq \left(\frac{1}{4}\right)\frac{(n-1)^2}{h_{n-1}} + \left(\frac{1}{2}\right)\frac{n}{h_n} \\
&= \left(\frac{1}{4}\right)\frac{(n-1)^2\frac{1}{n} + (n-1)^2 h_{n-1} + 2nh_{n-1}}{h_n h_{n-1}} \\
&> \left(\frac{1}{4}\right)\frac{n^2+1}{h_n} > \left(\frac{1}{4}\right)\frac{n^2}{h_n}
\end{aligned}$$

Thus $g_n = \Omega(n^2/\log n)$, as desired. ∎

In the course of the proof above we established that $\triangle g_{n-1} = \sum_{k=0}^{n} f_k \geq (1/2)(n/h_n)$. Using an argument similar to that of Lemma 3.4 one can prove that $\triangle g_n = O(n/\log n)$. Thus an immediate corollary to Lemma A.4 is

Corollary A.5

$$\sum_{k=0}^{n} f_k = \theta\left(\frac{n}{\log n}\right).$$

Bibliography

[1] D.J. Aldous. "On the Time Taken by Random Walks on Finite Groups to Visit Every State". *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, pp. 361-374, **62** (1983).

[2] D.J. Aldous. "Reversible Markov Chains and Random Walks on Graphs". Draft of first six chapters of book, January 26, 1993.

[3] R. Aleliunas, R. M. Karp, R. J. Lipton, L. Lovász, and C. Rackoff. "Random Walks, universal traversal sequences, and the complexity of maze problems". *20th Annual Symposium on Foundations of Computer Science*, pp. 218-223, San Juan, Puerto Rico, October 1979.

[4] G. Barnes and U. Fiege. "Short Random Walks on Graphs". *Proceedings of the 25th Annual ACM Symposium on Theory of Computing*, pp. 728-737, May 1993.

[5] J.A. Bondy and U.S.R. Murty. *Graph Theory with Applications.* New York: Elsevier North Holland, Inc., 1976.

[6] G. Brightwell and P. Winkler. "Maximum hitting times for random walks on graphs". *Random Structures and Algorithms*, 1:263-276, 1990.

[7] G. Brightwell and P. Winkler. "Extremal Cover Times for Random Walks on Trees". *Journal of Graph Theory*, 14:5, pp. 547-554(1990).

[8] Paul S. Bruckman. private communication.

[9] A. Chandra, P. Raghavan, W. Ruzzo, R. Smolensky and P. Tiwari. "The Electrical Resistance of a Graph Captures its Commute and Cover Times". *Proceedings of the 21st Annual ACM Symposium on Theory of Computing*, pp. 574-586, May 1989.

[10] D. Coppersmith, U. Feige and J. Shearer. "Random Walks on Regular and Irregular Graphs". *Preprint*, September 22, 1993.

[11] P.G. Doyle and J.L. Snell. *Random Walks and Electrical Networks*, The Mathematical Association of America, 1984.

[12] William Feller. *An Introduction to Probability Theory and Its Applications*, Vol 1, John Wiley and Sons, Inc. 1957.

[13] Uriel Fiege. "A Tight Upper Bound on the Cover Time for Random Walks on Graphs". *Random Structures and Algorithms*, Vol 6, No. 1, 1995.

[14] Uriel Fiege. "A Tight Lower Bound on the Cover Time for Random Walks on Graphs". *Preprint*, October 21, 1993.

[15] Daniel C. Fielder. private communication.

[16] F. T. Howard. "Congruences and Recurrences for Bernoulli Numbers of Higher Order". *The Fibonacci Quarterly*, Vol 32, No. 4, pp. 316-328, 1994.

[17] Charles Jordan. *Calculus of Finite Differences.* New York: Chelsea, 1965.

[18] J.G. Kemeny and J.L. Snell. *Finite Markov Chains*, Springer-Verlag, 1983.

[19] László Lovász and Peter Winkler. "Note: On The Last New Vertex Visited by a Random Walk". *DIMACS Technical Report*, 91-38, May 1991.

[20] Peter Matthews. "Covering Problems for Brownian Motion of Spheres". *The Annals of Probability*, Vol 16, No. 1, pp. 189-199, 1988.

[21] Niels Erik Nörlund. *Vorlesungen Über Differenzenrechnung.* New York, Chelsea, 1954.

[22] Prasad Tetali. "Random Walks and the Effective Resistance of Networks". *Journal of Theoretical Probability*, 4:101-109, 1991.

[23] Herbert S. Wilf. "The Editor's Corner: The White Screen Problem". *MAA Monthly*, Vol 96, No 8, pp. 704-707, Oct '89.

[24] P. Winkler. "Extremal Problems for Random Walks on Graphs". *Preprint.*

[25] P. Winkler and D. Zuckerman. "Multiple Cover Time". Unpublished Extended Abstract.

[26] David Zuckerman. "Covering Times of Random Walks on Bounded Degree Trees and Other Graphs". *The Journal of Theoretical Probability*, Vol 2, No 1, pp. 147-157, Jan '89.

[27] David Zuckerman. "On the time to traverse all edges of a graph". *Information Processing Letters*, 38:335-337, 1991.

[28] David Zuckerman. "A Technique for Lower Bounding the Cover Time". *SIAM Journal of Discrete Mathematics*, Vol. 5, No. 1, pp. 81-87, Feb '92.

Vita

Eric Richard Bussian was born on February 26th, 1958 in Summit, N.J.

He graduated from Charleston Catholic High School, Charleston, W.V. in 1976 and accepted an appointment to the United States Air Force Academy. He was awarded his Bachelor of Science degree and commissioned a Second Lieutenant in the United States Air Force in May 1980. He was then posted to Columbus Air Force Base, Columbus, Ms. to enter USAF pilot training.

In July 1981 2Lt. Bussian completed Undergraduate Pilot Training and was assigned to Sheppard AFB, Wichita Falls, Tx. as an instructor pilot in the Euro-NATO Joint Jet Pilot Training program: a program designed to provide primary flight training to most of the pilots from the NATO air forces. His instructional abilities were recognized when he was named the number one graduate of his pilot instructor training class.

After completing his tour of duty at Sheppard AFB, Captain Bussian was reassigned to Clark Air Base, Republic of the Philippines, as a cargo pilot. He flew the C-130E/H aircraft and became a Command Post Operations Officer.

Vita

Eric Richard Bussian [redacted]

He graduated from Charleston Catholic High School, Charleston, W.V. in 1976 and accepted an appointment to the United States Air Force Academy. He was awarded his Bachelor of Science degree and commissioned a Second Lieutenant in the United States Air Force in May 1980. He was then posted to Columbus Air Force Base, Columbus, Ms. to enter USAF pilot training.

In July 1981 2Lt. Bussian completed Undergraduate Pilot Training and was assigned to Sheppard AFB, Wichita Falls, Tx. as an instructor pilot in the Euro-NATO Joint Jet Pilot Training program: a program designed to provide primary flight training to most of the pilots from the NATO air forces. His instructional abilities were recognized when he was named the number one graduate of his pilot instructor training class.

After completing his tour of duty at Sheppard AFB, Captain Bussian was reassigned to Clark Air Base, Republic of the Philippines, as a cargo pilot. He flew the C-130E/H aircraft and became a Command Post Operations Officer.